The American Economic Review

ARTICLES

MARCH 1998

Shorter Papers

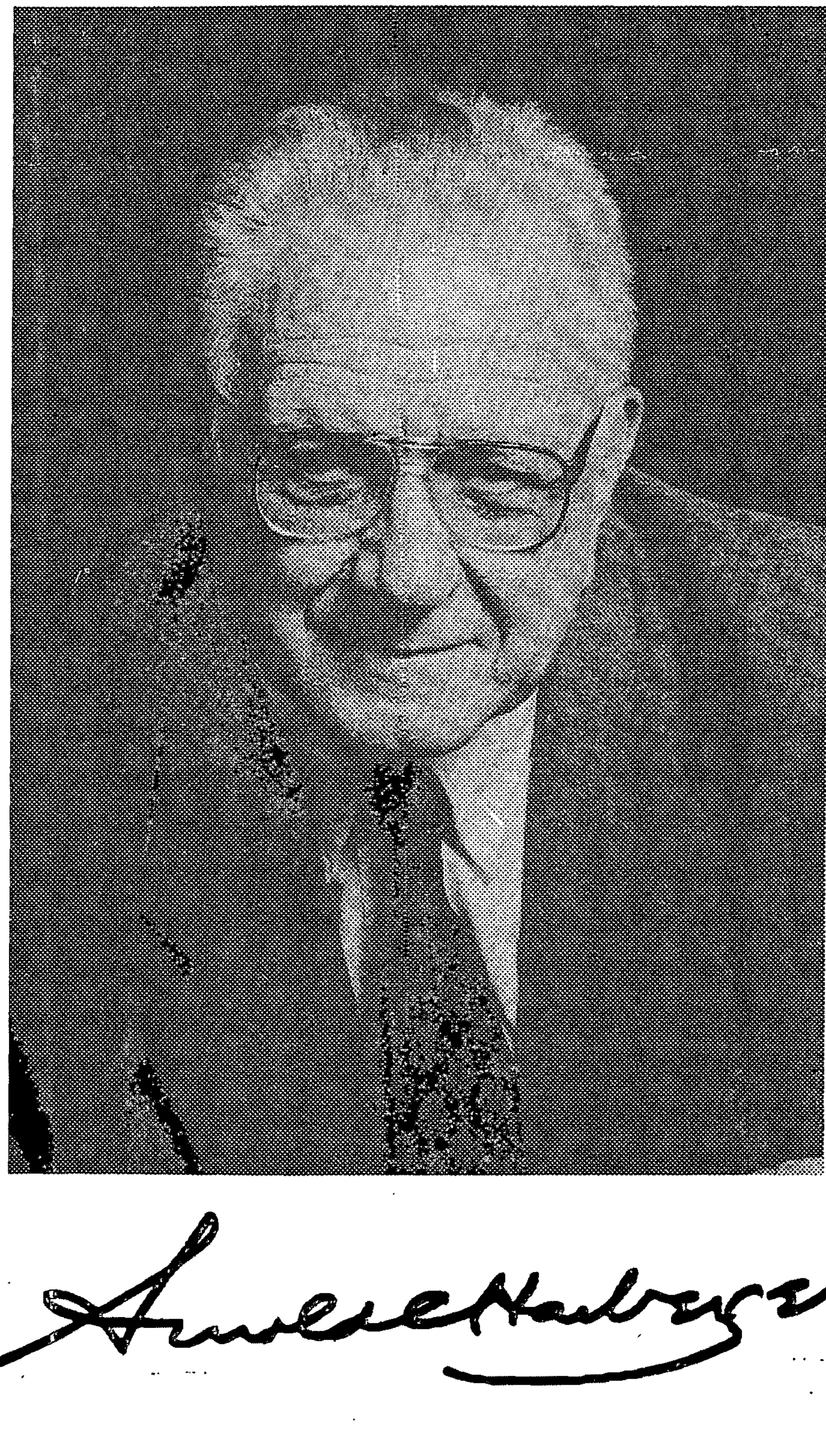
- Submit manuscripts (4 copies), single-sided, double-spaced, to:
  Orley Ashenfelter, Editor, *AER*, 209 Nassau Street,
  Princeton, NJ 08542-4607.

- Authorship should be identified only on a removable cover page; the anonymous text should begin on the following page.

- Submission fee: $50 for members; $100 for nonmembers. Please pay with a check or money order payable in United States dollars. Foreign (including Canadian) payments must be in the form of a check drawn on a United States bank payable in United States dollars.

- Style guides will be provided upon request.

It is the policy of the *American Economic Review* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary, or if, for some other reason, the above requirements cannot be met.



A Vision of the Growth Process†

By ARNOLD C. HARBERGER*

One of the great pleasures of belonging to y generation of economists is that we were able to witness the birth and the subsequent evolution of the modern approach to the analysis of economic growth. The centerpiece of that approach is probably growth accounting, but we should never forget that growth accounting is firmly rooted in economic theory.

My way of telling the story goes like this: Many, maybe even most, economists expected that increments of output would be explained by increments of inputs, but when we took our best shot we found that traditional inputs typically fell far short of explaining the observed output growth. Our best shot consisted in attributing to each factor a marginal product measured by its economic reward. Thus:

$$(1)\quad \bar{p}\Delta y = \bar{w}\Delta L + (\bar{\rho} + \delta)\Delta K + R.$$

Here:

Δy = change in output (GDP);
ΔL = change in labor input;
$\bar{p}$ = initial general price level;
$\bar{w}$ = initial real wage;
$\bar{\rho}$ = initial real rate of return to capital;
δ = rate of real depreciation of capital;
ΔK = change in capital stock; and
R = "the residual" of growth unexplained by increases in traditional inputs.

Many economists are probably more familiar with a variant of (1)

$$(1')\quad (\Delta y/y) = (\bar{w}L/\bar{p}y)(\Delta L/L) + [(\bar{\rho} + \delta)K/\bar{p}y](\Delta K/K) + (R/y) = s_\ell(\Delta L/L) + s_k(\Delta K/K) + (R/y).$$

In whichever form, the measured residual typically accounted for an important fraction of the observed output growth, quite often half or more.

† Presidential Address delivered at the one-hundred th meeting of the American Economic Association, uary 4, 1998, Chicago, IL.

* Department of Economics, University of California, Angeles, CA 90095. This paper is but one step in a uence of writings and other presentations on the pros of growth. In this work I have been greatly helped by umber of research assistants—Luis Alvarado and rald Beyer in the early stages, Alfonso Guerra and gar Robles in the middle phase, and Enrique Flores and onardo Torre in the final stage, as this paper was being pared. Readers will note that this paper also draws on D. dissertation work done by Beyer, Robles, and Torre. the course of the evolution just referred to, I have made sentations at, and received valuable comments from, a mber of different forums—seminars at Clemson Unisity, Stanford University, Texas A&M University, and LA in the United States, plus several abroad—the holic University of Chile, the University of Chile and Centro de Estudios Público (CEP) in Santiago, the nter for Argentine Macroeconomic Studies (CEMA), Instituto Torcuato Di Tella, the Universidad de San drés, and the Instituto Superior de Economistas del Gorno (ISEG) in Argentina, as well as the Instituto Teclogico Autónomo de Mexico (ITAM) in Mexico. In dition, there were conference presentations organized Cornell University, the East-West Center (University Hawaii), the Association of Asian Economists (Kuala mpur, 1996), the Western Economic Association (Sele, 1997), the Econometric Society (Latin American etings, Santiago, 1997), and the Argentine Political onomy Association (Bahia Blanca, 1997).

I want to make special note of a conference on capital mation and economic growth, organized by Michael J. skin and held at the Hoover Institution in October 1997. is conference was attended by a veritable galaxy of seus students of growth, a fact which motivated me to ite a very long paper (Harberger, 1998) probing issues methodology in the measurement and analysis of wth, as well as the more substantive matters emphaed in the present paper. I take this opportunity to refer ders to that paper (forthcoming in a volume edited by skin) for more detailed treatment of methodological iss. I also want to give special thanks to Zvi Griliches, le W. Jorgenson, and Paul Romer, as well as to my lleagues Sebastian Edwards, Jean-Laurent Rosenthal, d Carlos Végh, each of whom gave that paper a very eful reading and provided me with extremely useful mments. Finally, I want to give special thanks to rraine Grams, who has performed miracles translating ndwritten scrawls faxed from four continents into the per you see before you.

This result came as a surprise to the profession, though perhaps less so to those who reached it, or something very like it, by an alternative route. They were the people who came at the problem out of a tradition of measuring labor productivity, and at some point complemented output per worker with a measure of output per unit of capital, and finally joined the two to create a measure of total factor productivity (TFP). The idea of total factor productivity increasing through time was less a shock to these people than the "growth residual" was to those who approached its measurement along the lines of equation (1) or (1′). See Moses Abramovitz (1952, 1956) and Solomon Fabricant (1954).

In any case, as the newly discovered residual loomed large in our professional thinking, our discussion centered on two potential explanations: "human capital" and "technical advance." (See Robert M. Solow, 1957.) These can be thought of as complementary explanations, at least up to a point, with technical advance representing truly new ways of doing things, and the accumulation of "human capital" representing increases in the "quality" of the typical human agent. It was not long before attempts were made to quantify the contribution of improved labor quality. These came as part of a general move toward disaggregation of the two factors, which can be represented by:

$$\bar{p}\Delta y = \Sigma_i \bar{w}_i \Delta L_i + \Sigma_j(\bar{\rho}_j + \delta_j)\Delta K_j + R'. \tag{2}$$

Here the index i can vary over all sorts of education and skill groups as well as categories like gender, age, occupation, region, etc. All these are items that may signal a different market wage. In a similar vein, the index j would appropriately vary over categories like the corporate, noncorporate, and housing sectors where, for tax if for no other reasons, different (gross-of-tax) rates of return would presumably prevail, even in a full equilibrium.

In an equation like (2), the presumed marginal product of each category of labor is measured by the wage $\bar{w}_i$. Average quality can be measured by $Q_t = \Sigma_i \bar{w}_i L_{it} / \Sigma_i L_{it}$, and the contribution of change in quality to Δy_t between t and $t+1$ can be calculated as $L_t \bar{w}_t (\Delta Q_{t+1} / Q_t)$. Thus, the contribution of quality change is already built into the first summation in (2) but can be separately identified if we so choose.

A focus on human capital could lead us to a slightly different way of breaking down $\Sigma w_i \Delta L_i$. Here we could choose some "base wage" w^*, ideally the wage of some well-defined category of relatively unskilled labor. Then we could divide the remuneration w_i of any given category into a part w^* which we call a reward for "raw labor" and another part $(w_i - w^*)$ which we would identify as the reward to the human capital of a typical worker of type i.

Using a framework like (2) has long been the standard for careful professionals. Pioneered by Zvi Griliches (1960, 1963), it was utilized by Edward F. Denison (1967) and John W. Kendrick (1973, 1976, 1977), among others. This approach has been further developed and carried to a high art by Dale W. Jorgenson and Griliches (1967), Jorgenson et al. (1987), and Jorgenson (1995).

The main point to be made here is that once the residual is measured using a framework like (2) or its equivalent, the direct, measured contribution of human capital is captured in the labor term $\Sigma \bar{w}_i \Delta L_i$. By direct contribution I mean what people are paid for. Doctors earn more than nurses, and engineers more than draftsmen. These and similar differences are captured in $\Sigma \bar{w}_i \Delta L_i$, which can be positive even if $\Sigma \Delta L_i$ is zero, just from an upward reshuffling of the same labor force. A truly accurate measurement of type (2) would capture all the subtle differences of quality that exist in a modern labor force and would give each a weight corresponding to the (gross-of-tax) earnings that demanders are observed to pay. We may do this imperfectly, but, in concept at least, the residual R' as measured by (2) does not contain any elements of quality change or any direct contributions of human capital to growth. This is a quite important point for it permits us to zero in on the residual as representing "technical change," "TFP improvement," and "real cost reduction."

There is no analytical reason to prefer one of the above three terms over another, in referring to the residual R'. But I am going a bit out on a limb to say that a term like "technical change" leads most economists to think of inventions, of the products of research and de-

development (R&D), and of what we might call technical innovations. On the other hand, TFP improvement, once purged of the changes in the quality of labor and/or the direct contributions of human capital, makes one think of externalities of different kinds—economies of scale, spillovers, and systematic complementarities. And finally, real cost reduction, to my mind, makes one think like an entrepreneur or CEO, or a production manager.

I think it would be perfectly fair to characterize my presentation today as a paean in praise of "real cost reduction" as a standard label for R'. Labels do not change the underlying reality, but they may change the way we look at it and the way we think about it. They also can lead us to understand it better. Thinking in terms of real cost reduction has certainly done all this for me, as I have tried to sort out the many puzzles and complexities that surround the process of economic growth.

Let me try to take you down the path I traveled. In the first place, real cost reduction (RCR) is probably on the mind of most business executives, production managers, etc., at some point or another in any given week, let alone in any given month or year. It is a major path to profit in good times, and a major defense against adversity in bad times. Most U.S. firms that have downsized in recent years did so with RCR in mind. So, too, did the firms that computerized their payrolls and other accounts. And so also did those who shifted to what they considered more modern management techniques. I recall going through a clothing plant in Central America, where the owner informed me of a 20-percent reduction in real costs, following upon his installation of background music that played as the seamstresses worked. And then there is the story of two Chilean refrigerator firms that ended up as parts of a single conglomerate at one point. The new management reduced the number of models from something like 24 to two, making agreements to import other models while exporting these two. The end result was that output more than doubled, while the labor force was cut to less than half, and even the capital stock (at replacement cost) was significantly reduced. This sounds like (and is really) economies of scale, but they would not be detected by our usual measures, as both labor force and capital stock went down. And we all have seen cases where, say, an office's real costs were reduced when a martinet of a manager was replaced by someone more reasonable. But we have also seen cases where real costs were reduced when a very lax manager was replaced by someone more strict.

It has long been my song that there are at least 1001 ways to reduce real costs and that most of them are actually followed in one part or other of any modern complex economy, over any plausible period (say, a decade). Once one accepts this proposition as true, the question then arises: Why would anybody try to settle on just one underlying cause of real cost reduction? The answer, I think, is *mind-set*—the framework in which one is thinking at the moment. The pioneer writings of the recent endogenous growth literature can, I think, be said to reflect a kind of annoyance at something like R or R' being considered exogenous. There was an urge to surmount that inelegance by somehow making the residual endogenous. And in a simple growth model that meant generating a feedback from the rest of the model to the residual. A 1001 feedbacks would be out of the question, but one feedback would work just fine. Thus Paul Romer (1986) focused on a feedback through "knowledge," with the stock of knowledge shifting production functions all over the economy; Robert E. Lucas, Jr. (1988) focused on "human capital," not on its direct and remunerated productivity, but on the externalities that each increase in the stock of human capital were presumed to generate. These single feedbacks achieved the limited purpose of endogenizing R or R' within a specified model, but they did not represent very well the multifaceted nature of real cost reduction as we observe it in actuality. And, in point of fact, both the cited authors in their more recent writings display a deep recognition of the subtlety and complexity of the growth process, not really capable of being captured through a simple feedback mechanism. (See Romer, 1990, 1994a, b; Lucas, 1993.)

So, real cost reduction is multifaceted and everywhere around us. Where does that get us? Or how can we get anywhere in the face of such complexity? The next step is to recognize that in spite of its complexity, real cost reduction can be reduced to a single metric, and can

TABLE 1—GROWTH BREAKDOWN TREATING REAL COST REDUCTION AS ADDITIVE

Industry	TFP growth over period (1.0 = 100 percent)	Absolute amount of real cost reduction [(1) × (4)]	Cum. sum of (2)	Initial value added	Cum. sum of (4)
	(1)	(2)	(3)	(4)	(5)
1	0.800	$80b.	$80b.	$100b.	$100b.
2	0.600	$120b.	$200b.	$200b.	$300b.
3	0.500	$150b.	$350b.	$300b.	$600b.
All the rest	0.107	$150b.	$500b.	$1,400b.	$2,000b.

be made *additive.* For a quick appreciation of this, assume that total factor productivity grew by 80 percent in one industry over a decade, by 60 percent in another industry, and by 50 percent in a third. If their initial value added amounted to $100 billion, $200 billion, and $300 billion, respectively, then the real cost reduction of the first was $80 billion, that of the second was $120 billion, and that of the third $150 billion. So we can say that, measured at initial prices, the real cost reduction of the three together was $350 billion over the decade in question. I truly think that the notion of real cost reduction being additive in this way came to my mind, and is easily seen by others, just as a consequence of the label. The idea of additivity does not follow nearly so easily from the labels "technical advance" and "total factor productivity."

Anyway, this vision of the growth process opens up many new vistas and gives us many new challenges. To me, it gives *life* to the residual, viewed as real cost reduction, in a way that remote macroeconomic externalities never did. It gives the residual *body,* in the sense that the number of dollars saved by real cost reduction is a tangible and measurable quantity. It gives the residual a *name* (real cost reduction), an *address* (the firm), and a *face* (the face of the entrepreneur, the CEO, the production manager, etc.) And, finally, we shall see that there can be vastly different *expressions* on that face, even as we move from firm to firm in a given industry, as the TFP experience of a period moves from sharply positive to devastatingly negative.

I. Yeast versus Mushrooms: Part I

Table 1 is based on the numerical example just given, plus the information that the remaining industries (say, in the economy) together had an initial value added of $1,400 billion and experienced real cost reduction of $150 billion over the period. Setting out data in the format of Table 1 allows us to make statements like "15 percent ($300 b./$2,000 b.) of the industries (measured by initial value added) accounted for 40 percent ($200 b./$500 b.) of the real cost reduction (RCR) of the period" and "30 percent ($600 b./$2,000 b.) of the industries accounted for 70 percent ($350 b./$500 b.) of the period's RCR."

I stumbled on this way of presenting data on real cost reduction in the course of writing a background paper (Harberger, 1990) for the World Bank's *World Development Report* of 1991. Once I saw it, I immediately embraced it, because it helped me communicate to others what I call the "yeast versus mushrooms" issue. The analogy with yeast and mushrooms comes from the fact that yeast causes bread to expand very evenly, like a balloon being filled with air, while mushrooms have the habit of popping up, almost overnight, in a fashion that is not easy to predict. I believe that a "yeast" process fits best with very broad and general externalities, like externalities linked to the growth of the total stock of knowledge or of human capital, or brought about by economies of scale tied to the scale of the economy as a whole. A "mushroom" process fits more readily with a vision such as ours, of real cost

TABLE 2—CONCENTRATION OF TFP GROWTH AMONG U.S. INDUSTRIES 1958–1967
[COLUMNS (2) TO (5) IN BILLIONS OF 1958 DOLLARS]

	TFP growth over period (1.0 = 100 percent)	Absolute amount of real cost reduction [(1) × (4)]	Cum. sum of (2)	GDP by industry 1958	Cum. sum of (4)
	(1)	(2)	(3)	(4)	(5)
umber & Wood Products	0.72	2.51	2.51	3.50	3.50
ailroad Transport	0.63	5.52	8.03	8.70	12.20
extile Mill Products	0.61	2.49	10.52	4.10	16.30
lectrical Machinery	0.55	5.10	15.66	9.30	25.60
ransport Equipment	0.46	7.05	22.71	15.20	40.80
hemicals	0.44	3.97	26.68	9.10	49.90
ublic Utilities	0.42	4.65	31.33	11.00	60.90
troleum and Coal	0.41	1.27	32.60	3.10	64.00
ubber and Products	0.41	1.23	33.83	3.00	67.00
ining	0.41	5.20	39.03	12.60	79.60
ommunication	0.40	3.61	42.64	9.00	88.60
rade	0.33	24.93	67.57	76.40	165.00
here follow 18 more industries e combined results of which are	0.03	7.53	75.10	239.80	404.80

otes: Top 10 percent (these percentages are contributions to GDP of industries ranked according to their present rate of FP growth over period) of industries account for 30 percent of total TFP contribution.
p 22 percent (these percentages are contributions to GDP of industries ranked according to their percent rate of TFP owth over period) of industries account for 52 percent of total TFP contribution.
p 40 percent (these percentages are contributions to GDP of industries ranked according to their percent rate of TFP owth over period) of industries account for 70 percent of total TFP contribution.
urces: Kendrick and Grossman (1980). GDP data from U.S. national accounts.

ductions stemming from 1001 different uses, though I recognize that one can build enarios in which even 1001 causes could rk rather evenly over the whole economy. Personally, I have always gravitated toward e "mushrooms" side of this dichotomy. I member being impressed, when I first saw me early industry estimates of TFP improvement, by their tendency to industry concentran. For years I told my students that the 20's were the decade of cars and rubber es, the 1930's the decade of refrigerators, e 1940's that of pharmaceuticals (especially tibiotics), and the 1950's that of television, with telecommunications and computers taking over in recent decades. But these were just impressions, not based on any systematic approach. My real turnaround came in the course of writing my 1990 paper, where I presented a series of tables based on Kendrick and Elliot S. Grossman's (1980) work. Table 2 is an example.

Table 2 has the same format as Table 1. Column (1) presents the familiar measure of the percentage by which TFP grew, or real costs were reduced, during the period in question (note that the percentages apply to the period 1958–1967 as a whole; they are not annual

rates). To turn these percentages into dollar amounts of real cost saving over the period, one multiplies them by base-period real GDP [col. (4)]. The results are shown in column (2). Columns (3) and (5) are the cumulative sums of columns (2) and (4), respectively. Working with these figures one can make statements like those at the bottom of the table—i.e., the top 10 percent of industries accounted for 30 percent of total real cost reduction; the top 22 percent of industries (measured by initial value added) accounted for more than half of total real cost reduction.

Readers will notice that at the foot of each column in the table is an entry referring to 18 additional industries, which together accounted for only 10 percent of the total TFP contribution, while their combined share of initial output was almost 60 percent of the total.

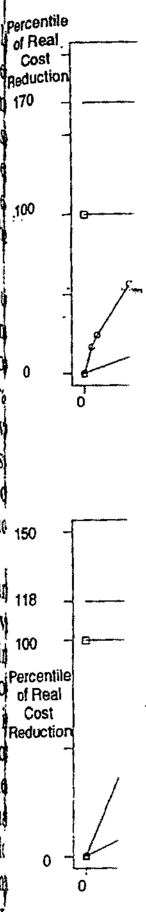
Using the analogy with yeast and mushrooms, the results of my calculations using the Kendrick-Grossman data pointed very clearly to a "mushrooms" interpretation. Not only were the contributions to RCR highly concentrated in a relatively few industries, these industries also were very different as one shifted from decade span to decade span. The top four branches in percentage of real cost reduction during 1948–1958 were Communications, Public Utilities, Farming, and Miscellaneous Manufacturing. In 1958–1967 they were Lumber, Railroad Transport, Textile Mills, and Electrical Machinery. In 1967–1976 they were Finance, Insurance & Real Estate, Apparel, Communications, and Chemicals. Only Communications appears twice among these 12 listings.

Now to my mind, this already brings evidence to bear on a number of possible hypotheses concerning the nature of TFP improvement. Certainly some ways of interpreting a generalized externality due to improved education would be hard to justify using evidence like this. Strong links of the residual term to R&D expenditures[1] would suggest a high degree of persistence among the leaders in TFP improvement. So also (probably) would economies of scale associated with the scale either of the firm or of the industry.

Such economies are not likely to jump wild around from one industry to the next, from p riod to period. One would expect them to en body characteristics of the productive proce that would be relatively stable over tim hence they should show a reasonably high d gree of persistence, over time, in terms of t TFP experience of particular industries.

No economist can look at Table 2 witho thinking of its close analogy with a Lore curve. That, indeed, was the next step I too in trying to represent the degree of concentr tion of real cost reduction. Figure 1 (draw from Edgar Robles, 1997), shows the quas Lorenz curves for a 20-industry breakdown the U.S. manufacturing sector over four su cessive five-year periods.

What strikes one immediately about Figu 1 is the characteristic "overshooting." I ha marked with the first vertical line the poi where the rising curve crosses 100 percent the vertical axis. The interpretation is that 1970–1975 the cumulative real cost reducti of just 25 percent of manufacturing industri (measured by initial value added) was equ to the total RCR for manufacturing as a who After that there are other industries produci another 40 percent of the total, but their co tribution is offset by still other industries wi negative RCR during the period.



Corresponding to the 25-percent figure f 1975, we have around 12 percent for 1975 1980, 48 percent for 1980–1985, and 40 pe cent for 1985–1991. These are the fractions manufacturing industry which by themselv were able to account for the full amount of re cost reduction during the respective period, manufacturing industry as a whole.

The second vertical line in each panel Figure 1 marks the maximum point of t curve. The interpretation is that about 64 pe cent of industries enjoyed real cost reducti during 1970–1975, with the remaining 36 pe cent suffering real cost increases (declini TFP). For the subsequent periods, the corr sponding figures are 65(35) percent, 78(2 percent, and 82(18) percent. Here the first fi ure is the percent of industries enjoying re cost reductions; the figures in parentheses re resent those experiencing declining TFP.

Some interest attaches to the ordinate of t maximum point on each curve. In the first p riod, TFP growth ended up accounting for clo

[1] For a review of the current status of analysis of R&D expenditures and their impact on economic growth see Griliches (1994).

ıot likely to jump wil
stry to the next, from
ould expect them to e
f the productive proce
vely stable over tim
ıw a reasonably high d
ver time, in terms of t
rticular industries.
look at Table 2 with
analogy with a Lore
vas the next step I to
the degree of concent
ction. Figure 1 (dra
997), shows the qua
)-industry breakdown
ıg sector over four su
ıds.
ımediately about Figu
"overshooting." I ha
vertical line the poi
crosses 100 percent
interpretation is that
ative real cost reducti
nanufacturing industri
value added) was equ
anufacturing as a who
ıer industries produci
the total, but their co
till other industries wi
the period.
ıe 25-percent figure f
112 percent for 197
980–1985, and 40 pe
hese are the fractions
y which by themselv
ır the full amount of r
the respective period,
y as a whole.
l line in each panel
naximum point of t
on is that about 64 p
ıyed real cost reducti
th the remaining 36 p
st increases (declini
uent periods, the cor
55(35) percent, 78(2
ercent. Here the first fi
ndustries enjoying r
ures in parentheses r
ing declining TFP.
es to the ordinate of t
h curve. In the first p
l up accounting for cl

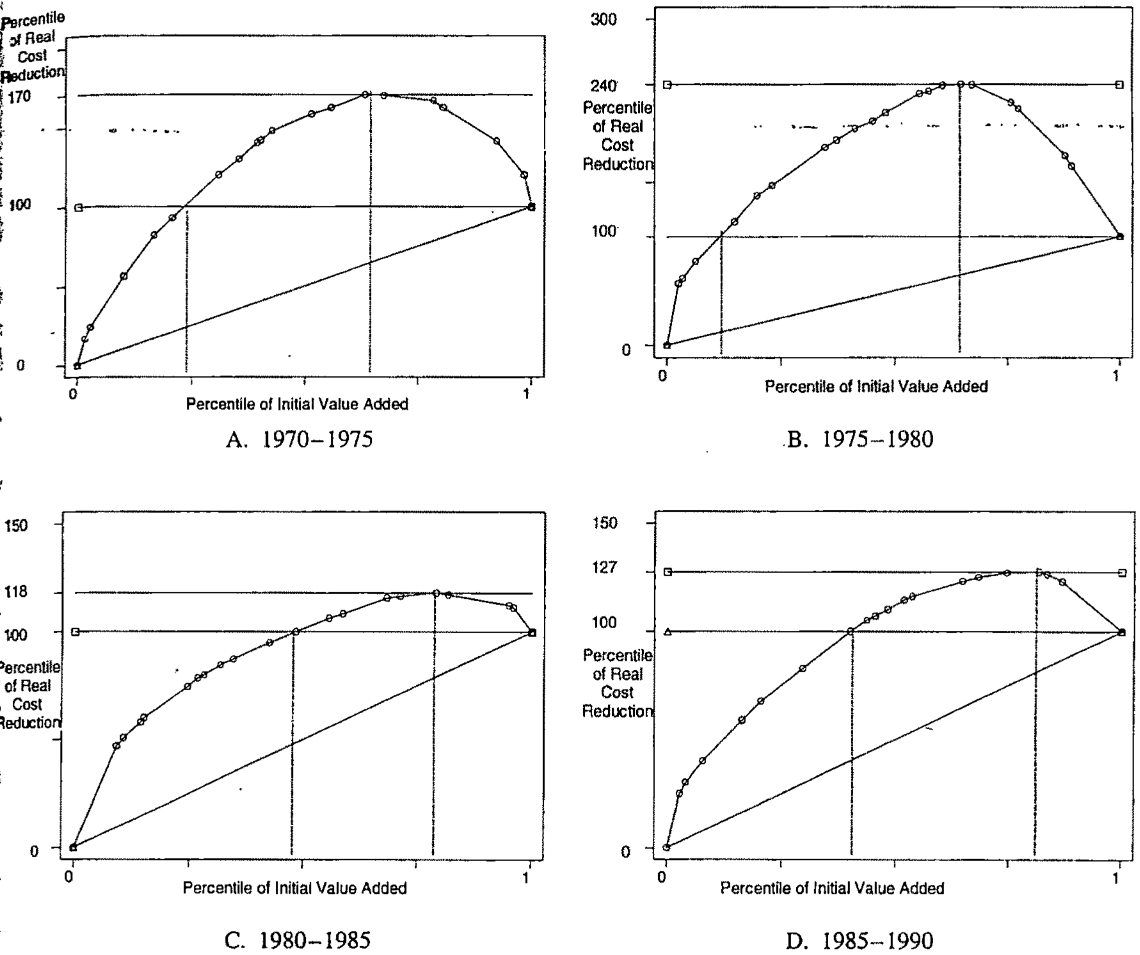


FIGURE 1. PROFILES OF TFP GROWTH AMONG U.S. MANUFACTURING BRANCHES

to 170 percent of the RCR for total manufacturing. In 1975–1980 this figure was about 240 percent, in 1980–1985 only about half that, and in 1985–1991 a little more than 125 percent. The trouble is that when the aggregate TFP contribution is relatively small, the cumulative total of the positive contributions is a large multiple of that aggregate, while when the aggregate is large, this multiple tends to be smaller. Thus, for 1970–1975 and for 1975–1980, the total RCR in manufacturing as a whole was only about 2.3 percent of initial manufacturing value added. In contrast, the total RCR for all manufacturing was almost 10 percent of initial manufacturing value added in 1980–1985, and about 7.5 percent in 1985–1991.

The problem obviously becomes greatly compounded if the real cost reduction for the aggregate (in this case total manufacturing) turns out to be negative. Special conventions would have to be established to make clear the interpretation of Lorenz-like diagrams in such cases.

I believe I have hit on a felicitous way of solving all these problems, and at the same time creating an even better, clearer visual representation of the degree of concentration or dispersion of real cost reduction among the components of an aggregate. The idea is simply to relabel the vertical axis of the Lorenz-like diagram, making it represent an annual growth rate. For simplicity, think of a 30-degree line as representing 1 percent per annum of TFP growth. The rest of the vertical axis would be calibrated accordingly. Thus, by looking at the slope of a simple chord, we could visually assess how rapid was the TFP growth of the aggregate in question.

Figure 2 is presented simply for didactic purposes. Here we have a hypothetical industrial branch made up of four industries, A, B,

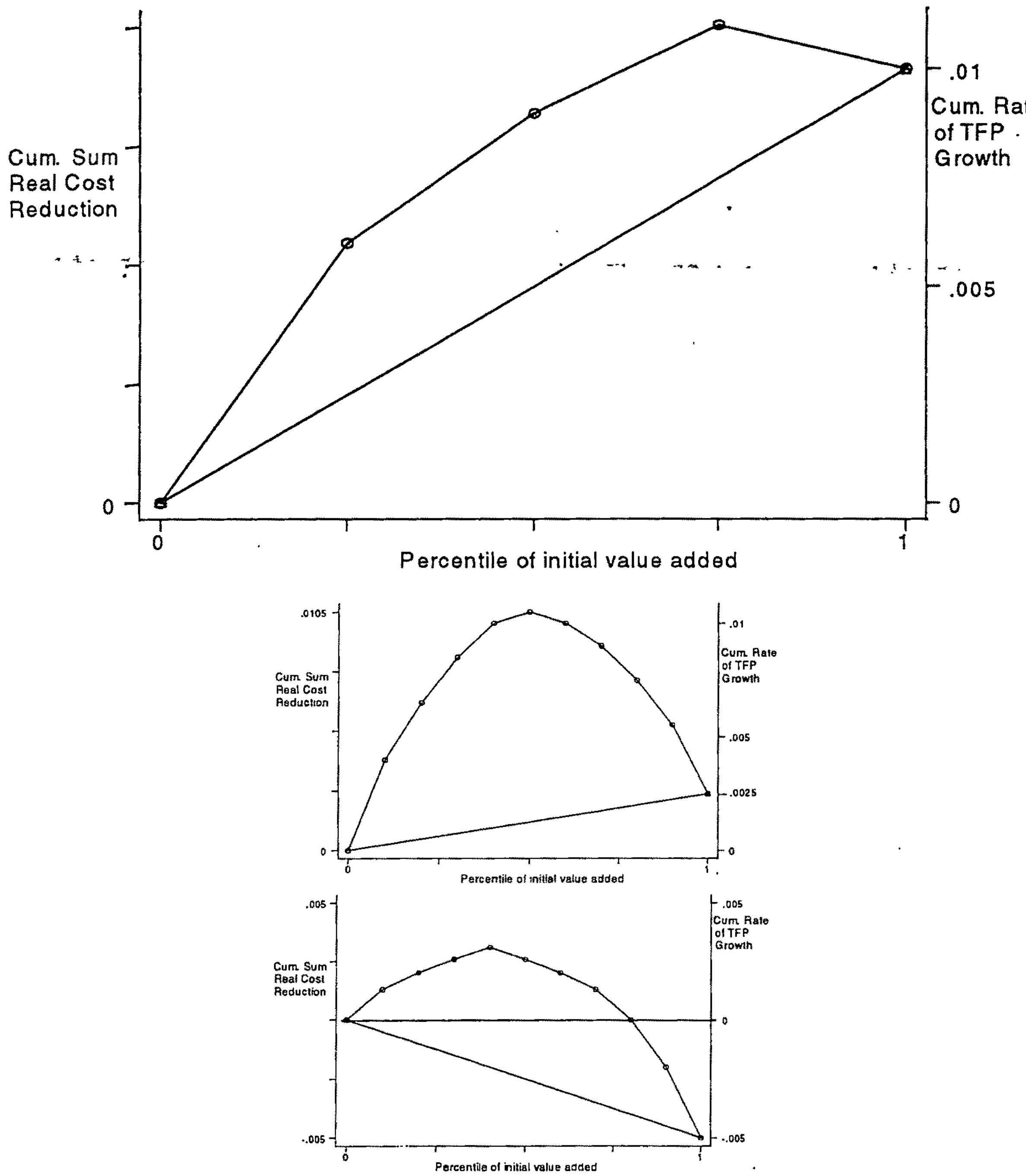


FIGURE 2. ILLUSTRATIVE TFP GROWTH PROFILES (SUNRISE-SUNSET DIAGRAMS)

C, and D. First, we order the industries in descending order, according to their rates of TFP increase in the period. Then we calculate cumulative real cost reduction (a real dollar amount) and plot it against cumulative initial real value added. Then we scale the vertical axis so as to comply with whatever metric we have decided upon for the TFP growth rate (in the example, a 30-degree line representing 1-percent annual TFP growth rate), and th horizontal axis so as to add up to 100 percen

In the lower panel of Figure 2 I give ex amples to show how these diagrams cope wit the problems of a low TFP growth rate (th overshoot for the case of 0.25-percent growt would show up peaking at over 400 percent i

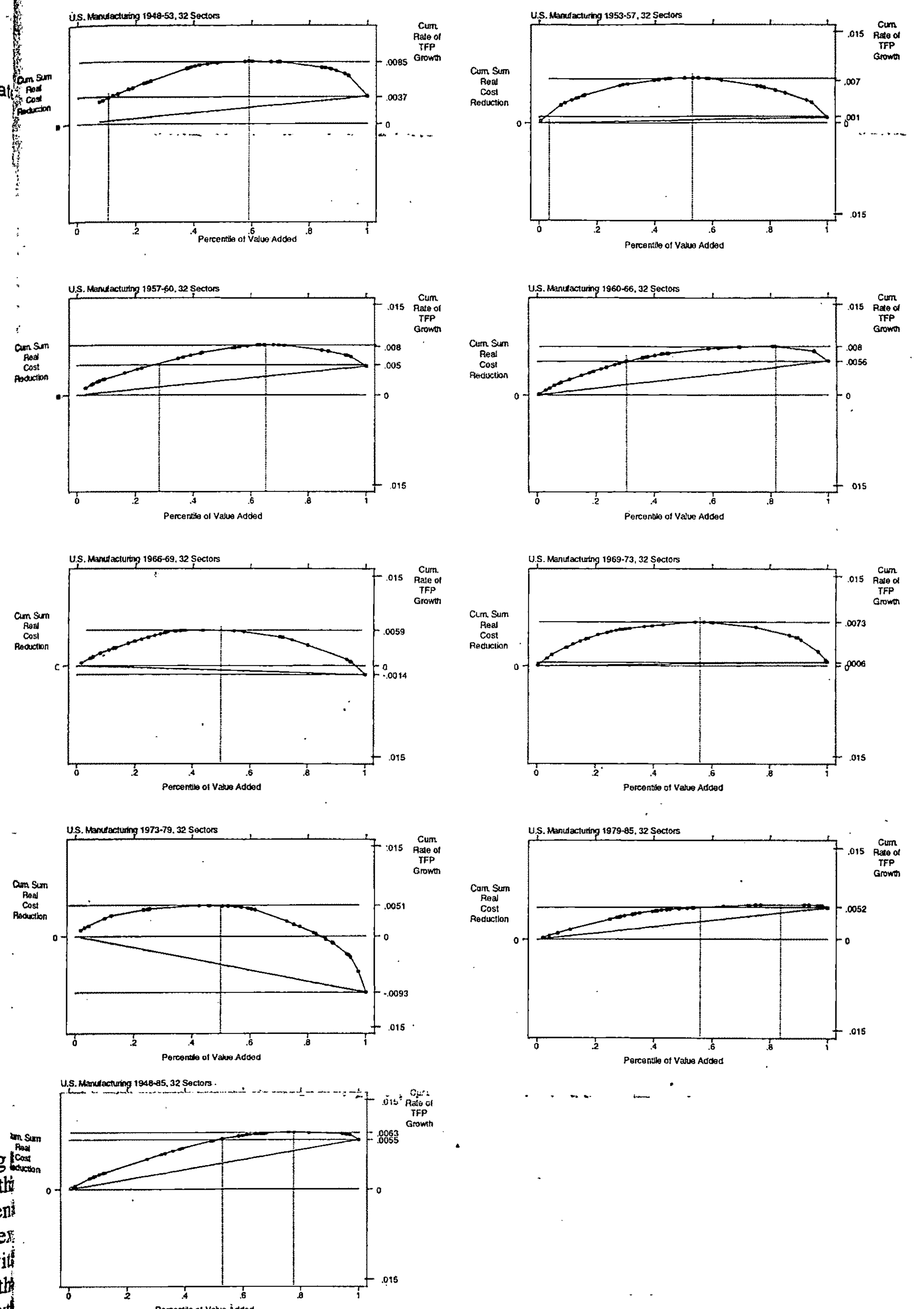


FIGURE 3. TFP GROWTH PROFILES FOR U.S. MANUFACTURING

a Lorenz-type diagram) and of negative TFP growth (where it is hard to even conceptualize a Lorenz-type picture).

I first presented these diagrams before a large audience at the Western Economic Association meetings in Seattle (July 1997), and for that presentation coined the label of "sunrise diagrams" on their analogy with the sun rising over a hill. That same evening Yoram Barzel suggested that where the aggregate slope is negative, we apply the term "sunset diagrams," which I immediately accepted.

Figure 3 presents a set of sunrise-sunset diagrams based on Jorgenson et al. (1987 pp. 188–90). These cover 32 industrial sectors (their 35 minus Agriculture, Trade, and Government Enterprises). I think the utility of sunrise-sunset diagrams needs no further championing once these pictures are examined and digested. Practically all variants are represented in these real-world cases: low TFP growth with a huge overshoot (1953–1957 and 1969–1973); negative growth with large and moderate overshoots (1966–1969 and 1973–1979); moderate growth with small (1979–1985), medium (1960–1966), and large (1948–1953) overshoots.

One striking fact that emerges from this set of pictures is how variable across periods is the negative contribution of the losers. If the losers had only contributed zero change in TFP, we would have had cumulative TFP contributions of about 0.8 percent per annum in 1948–1953, in 1957–1960, and in 1960–1966. And the other periods would not have been much different: about 0.7 percent in 1953–1957 and in 1969–1973, 0.6 percent in 1966–1969, and 0.5 percent in 1973–1979 and 1979–1985. Instead of this narrow range of cumulative contributions, we have an actual distribution that goes from −0.9 percent in 1973–1979 through around 0.1 percent in 1953–1957 and 1969–1973 to over 0.5 percent in 1960–1966 and 1979–1985.

Does this not suggest that we make a major research push trying to improve our understanding of the phenomenon of negative TFP growth? What syndromes characterize the firms and industries experiencing it? How much of it stems from external shocks like international prices? How much of it from competition within the industry? How much of it represents firms struggling to survive, yet experiencing output levels well below their previous peaks (and presumably below installed capacity)? How much of it represents things like "labor hoarding" as firms go through periods of adversity?

II. Yeast versus Mushrooms: Part II

I hope that in the previous section I have made a convincing case concerning: (a) the usefulness of sunrise-sunset diagrams, (b) the aptness of the "yeast versus mushrooms" dichotomy, and (c) the pervasiveness with which the mushroom side of that dichotomy seems to come out ahead when the GDP is broken down into industries or industrial branches for TFP analysis. The grand design that emerges from the studies reported here, and from just about all the other industry breakdowns that I recall having seen, is that: (i) a small-to-modest fraction of industries can account for 100 percent of aggregate real cost reduction in a period; (ii) the complementary fraction of industries contains winners and losers, the TFP contributions of which cancel each other; (iii) the losers are a very important part of the picture most of the time, and contribute greatly to the variations we observe in aggregate TFP performance; and (iv) there is little evidence of persistence from period to period of the leaders in TFP performance.

The above results are, I think, very interesting (in the sense of piquing our curiosity), very strong (in terms of their implications about the nature of the growth process), and very robust (in the sense that they have wide applicability over different data sets analyzed by different authors using at least somewhat different methods). But these results, so far, are quite compatible with what I might call an "industry view" of the TFP story. This is the way I, myself, looked at the growth process until quite recently—a vision that was reflected in my stories about rubber tires and autos in the 1920's, refrigerators and other household appliances in the 1930's, pharmaceuticals in the 1940's, etc. The image that I had in mind was one of yeast within each industry and mushrooms between industries—a commonality of TFP experience by firms within an industry, depending on that industry's

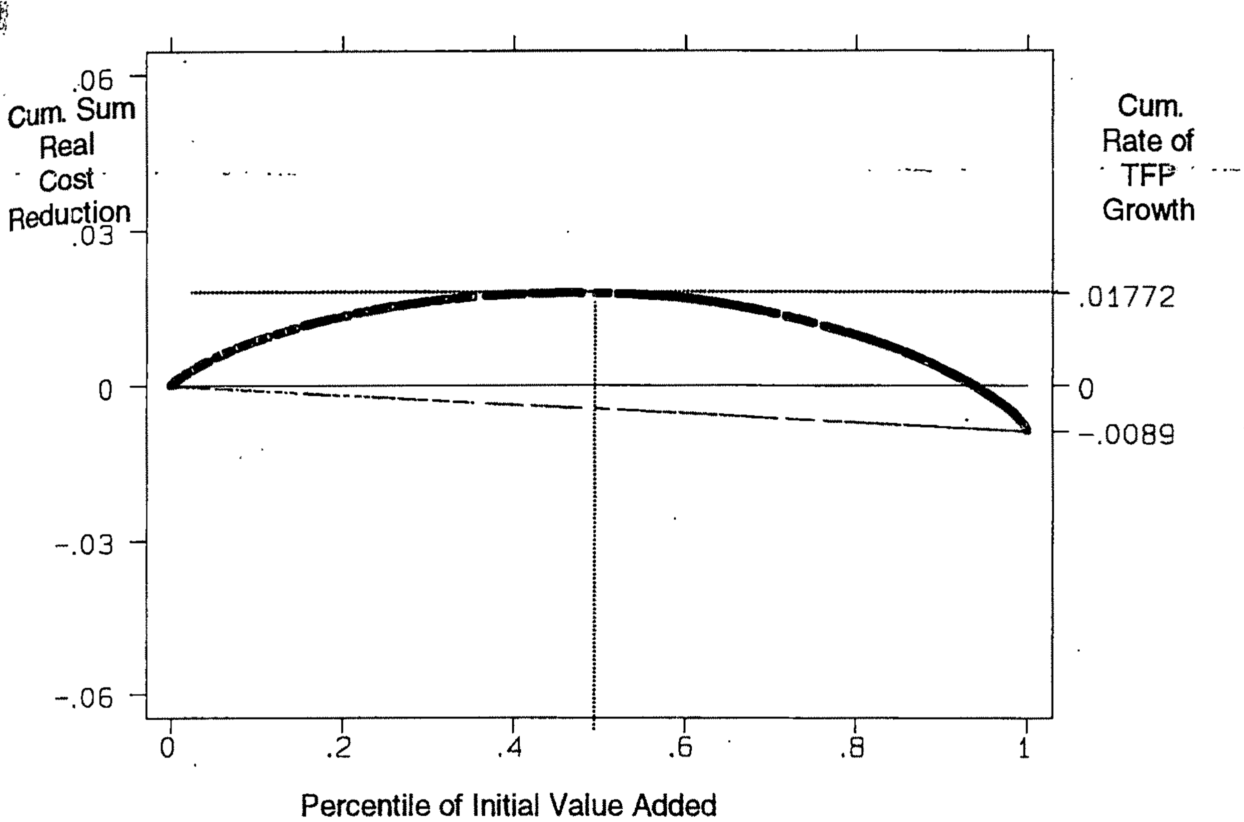


FIGURE 4. TFP GROWTH PROFILE IN MEXICAN MANUFACTURING SECTOR (1892 ESTABLISHMENTS, 1984–1994)

luck in the technological draw, side by side with highly diverse experience between industries because the distribution of technical advances had wide dispersion, even for periods as long as a decade.

Getting access to data at the firm level permits one to explore whether this view is compatible with the actual experiences of firms and industries. We are just in the early stages of this exploration, but I think the result is quite clear already; namely, the "mushrooms" story prevails just as much among firms within an industry as it does among industries within a sector or broader aggregate. I will present here only a taste of the evidence from the United States (on which our systematic work just recently got started). Our massive evidence comes from the Mexican manufacturing sector, for which Leonardo Torre (1997) has analyzed data from a sample of over 2,000 firms. A small fraction of these firms were lost owing to missing data, but some 1,900 firms remained in the sample that Torre finally worked with. These firms were divided into 44 branches of industry, so that on average we have about 43 firms per branch.

There are really too many ways to present such a mass of information as is contained in Torre's study. What I will do here is give the aggregate picture in Figure 4, and then show in Figures 5A–C three fast-growing branches, three of around median growth, and three from among the slowest-growing branches.

To complement these figures, I finally present, in Figures 6A–D, certain summary statistics from the sunrise-sunset diagrams of the 44 branches that Torre studied. Here Figure 6A gives the distribution of average rates of TFP growth among the 44 industries. Figure 6B shows the distribution of peak cumulative contributions, i.e., what the TFP contribution would have been had all the negatives been zeros. Figure 6C displays the percentile of firms (by initial value added) marking the borderline between positive and negative TFP growth. And finally, Figure 6D shows, for branches with positive TFP growth, the percentile of firms which, by themselves, account for 100 percent of the industry's TFP growth.

This evidence almost seems to replicate, for firms within an industry, what was found in the previous section for industries within the

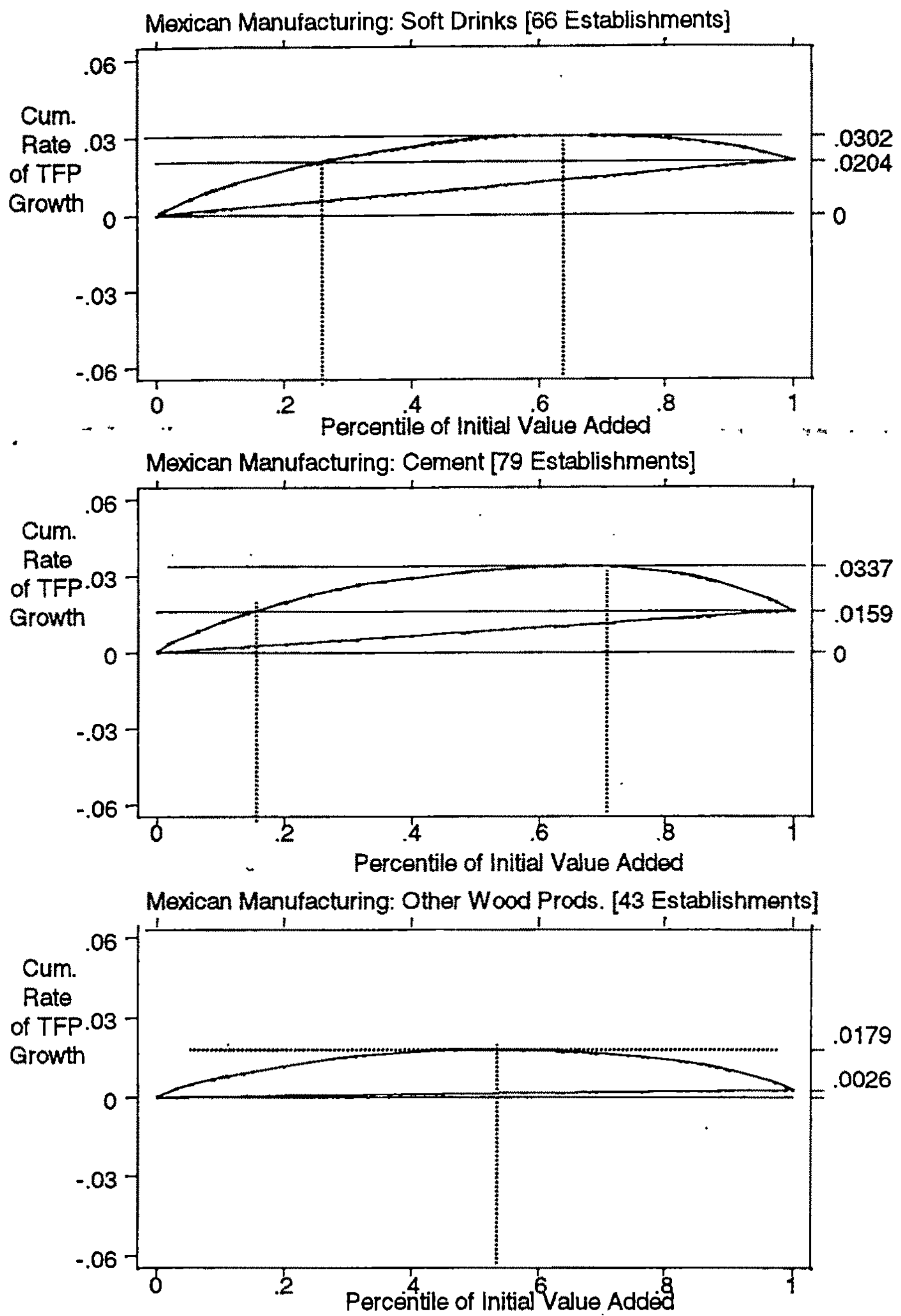
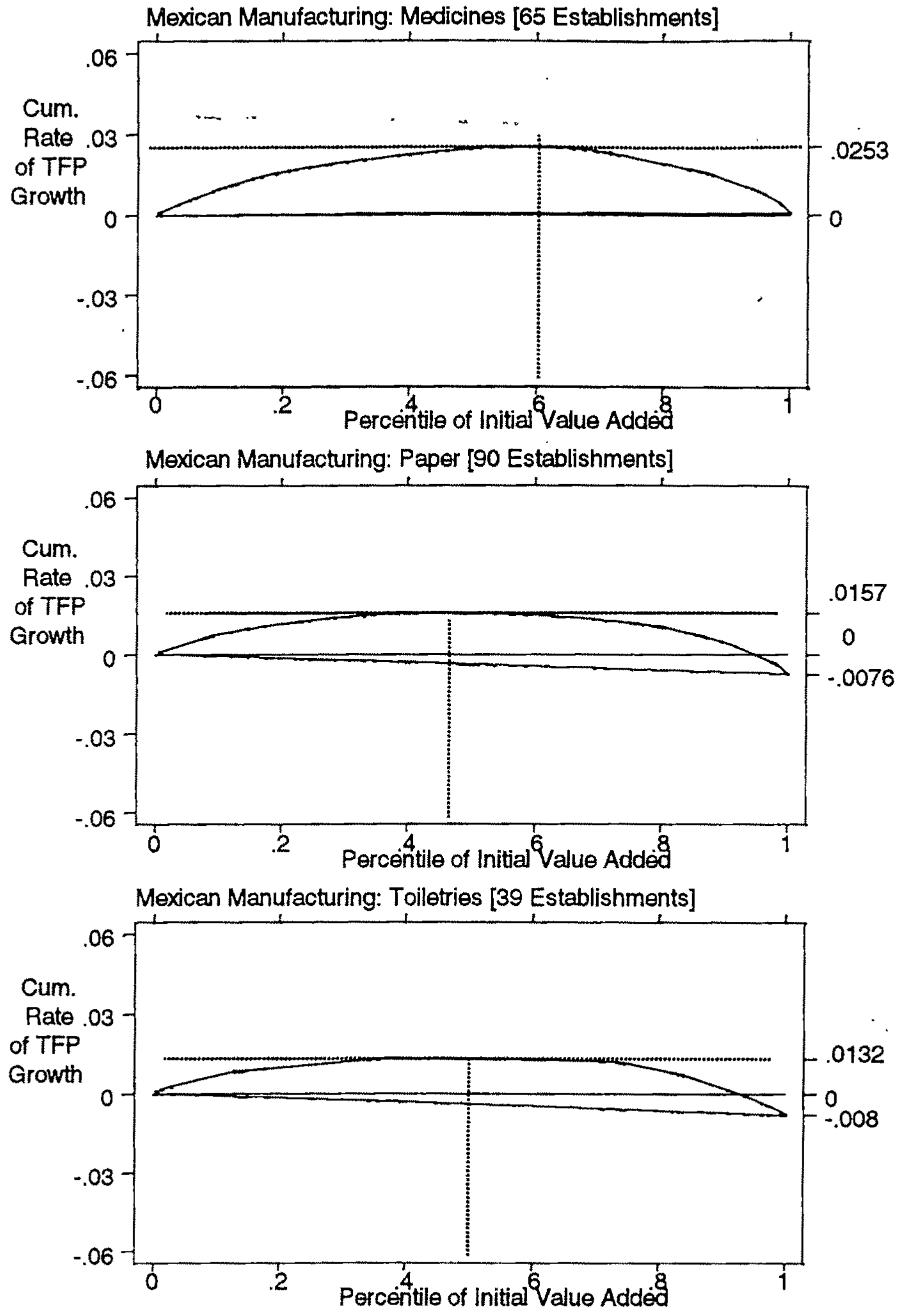


FIGURE 5A. TFP GROWTH PROFILES FOR FAST-GROWING BRANCHES (MEXICAN MANUFACTURING, 1984–1994)

economy—rampant overshooting of sunrise-sunset diagrams, great influence of firms with negative TFP growth in determining the TFP outcome for an industry, and a small or moderate fraction of firms accounting for 100 percent of the TFP growth of an industry (when that growth is positive), with the complementary fraction being winners and losers whose efforts end up just offsetting each other. It remains to try to give some interpretation to those results.

III. "Just Errors" or "It's a Jungle Out There?"

The first question that will enter the minds of many economists on looking at the evidence presented so far is: how much of what



;, 1984–1994) FIGURE 5B. TFP GROWTH PROFILES FOR MEDIUM-GROWING BRANCHES (MEXICAN MANUFACTURING, 1984–1994)

ve have seen and emphasized might simply
nterpretation toe the result of errors of observations? This
s by no means a frivolous question. For one
:an actually create frequency distributions
's a Jungle)f rates of TFP increase which contain ex-
ıctly the same information as the sunrise-
ıunset diagrams previously presented. The
enter the minonly trick is to count as the unit of frequency
:ing at the evilot one firm (out of an industry aggregate)
v much of whaor one industry (out of some larger aggre-
gate) but, instead, say, 1 percent of the total value added of the aggregate. Thus a firm with 20 percent of the value added of an industry would appear with 10 times the weight of a firm accounting for 2 percent of the value added of that industry. In such a chart, the cumulative frequency (say, 68 percent) above ΔTFP = 0 would represent the projection on the horizontal axis of the maximum point on a sunrise diagram. Its

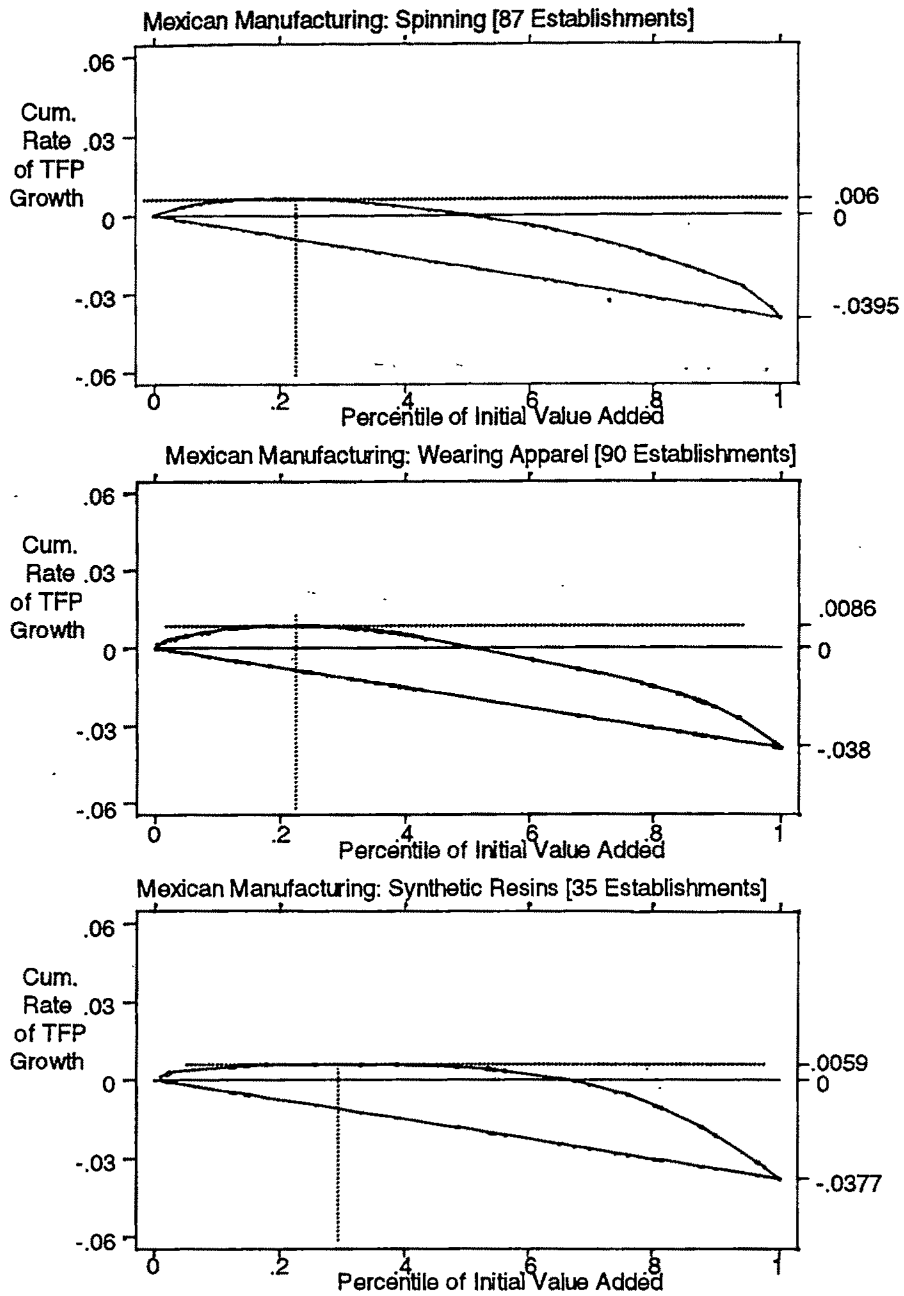


FIGURE 5C. TFP GROWTH PROFILES FOR SLOW-GROWING BRANCHES (MEXICAN MANUFACTURING, 1984–1994)

complement (32 percent) would represent the initial value added associated with negative TFP performance during the period.

If, then, all the information could be generated by a properly designed frequency distribution of rates of TFP growth, could it not all be the result of chance alone—more specifically, of errors of measurements? I really think not—my favorite quip on this is that "white noise does not sing a tune." That is, if we can rationalize what we see in terms of an analytical framework which embodies we established economic principles and sensit presumptions about underlying relationshi and facts, this is itself strong evidence agair the white noise hypothesis.

Nonetheless, we have to face the fact tl errors of observation of some magnitude ce tainly do exist, and we must recognize tl they can cloud our perceptions and bias c results. What I am going to do here is consic frequency distributions of firms. TFP is me

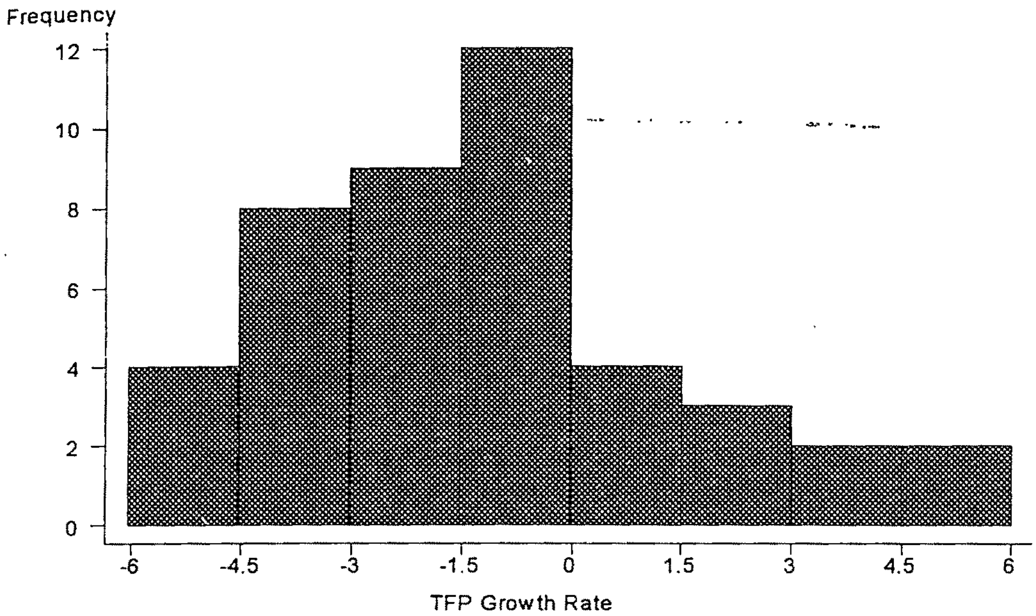


FIGURE 6A. AVERAGE ANNUAL TFP GROWTH RATE: MEXICAN MANUFACTURING 1984–1994 FREQUENCY DISTRIBUTION, 44 INDUSTRIAL BRANCHES

ured in two ways—one using value added by set of firms on the one hand, and the other sing "output" by those same firms, measured through dividing value added by separate estimated firm-by-firm price (of value dded) indexes p_j. For these purposes we can onveniently think in terms of *logarithms*, so et:

$_j$ = observed value added of firm;
$_j$ = estimated firm-level price index;
$_j = v_j - p_j$ = estimated output;
$_j = T_j + e_j$ [T_j = true value added];
$_j = \pi_j + u_j$ [π_j = true price index]; and
$_j = T_j - \pi_j$ [true output of firm].

NG, 1984–1994) Ve would like to have data on $\dot{q}_j$ and its ariance

$$\sigma_{\dot q}^2 = \sigma_{\dot T}^2 - 2\sigma_{\dot\pi\dot T} + \sigma_{\dot\pi}^2.$$

embodies well

les and sensib we simply work with observed value added ing relationships our quantity variable, we get evidence again

$$\sigma_{\dot v}^2 = \sigma_{\dot T}^2 + \sigma_{\dot e}^2 \text{ (assuming } \sigma_{\dot T\dot e} = 0\text{)}.$$

ace the fact th e magnitude ce we worked with the measured y_j, we get st recognize th ons and bias o

$$\sigma_{\dot y}^2 = \sigma_{\dot T}^2 + \sigma_{\dot e}^2 + \sigma_{\dot\pi}^2 + \sigma_{\dot u}^2 - 2\sigma_{\dot T\dot\pi}$$

here is consid ms. TFP is me (assuming $\dot u$ and $\dot e$ to be strictly random).

My presumptions are as follows:

(i) We can estimate value added quite accurately at the firm level. Hence the presumption that $\sigma_{\dot e}^2$ is small.

(ii) In most industries, there is considerable variety among the firms and their products. Hence, except in cases of industries with very homogeneous products, we should not expect $\sigma_{\dot\pi}^2$ to be small. Hence, I expect $\sigma_{\dot\pi}^2 > \sigma_{\dot e}^2$.

(iii) Finally, we have the presumption that, at least at the level of firms within an industry, $\sigma_{\dot\pi\dot T} < 0$. We know that firms choose to operate in regions of the demand curve where they consider the elasticity facing them to be greater than one. But also, in an analysis of the growth process, one would expect the big gains in value added to accrue to those firms in an industry which were passing along to consumers some of the fruits of current or past real cost reductions.

These three presumptions lead me to the conclusion that $\sigma_{\dot v}^2$ is likely to understate the true variance of output $\sigma_{\dot q}^2$ (because $-2\sigma_{\dot\pi\dot T} > 0$ and $\sigma_{\dot\pi}^2 > \sigma_{\dot e}^2$), and that $\sigma_{\dot y}^2$ is likely to overstate $\sigma_{\dot q}^2$ (only the covariance terms with $\dot e$ and $\dot u$, which were assumed to be zero, could make it otherwise). And since Torre worked with real

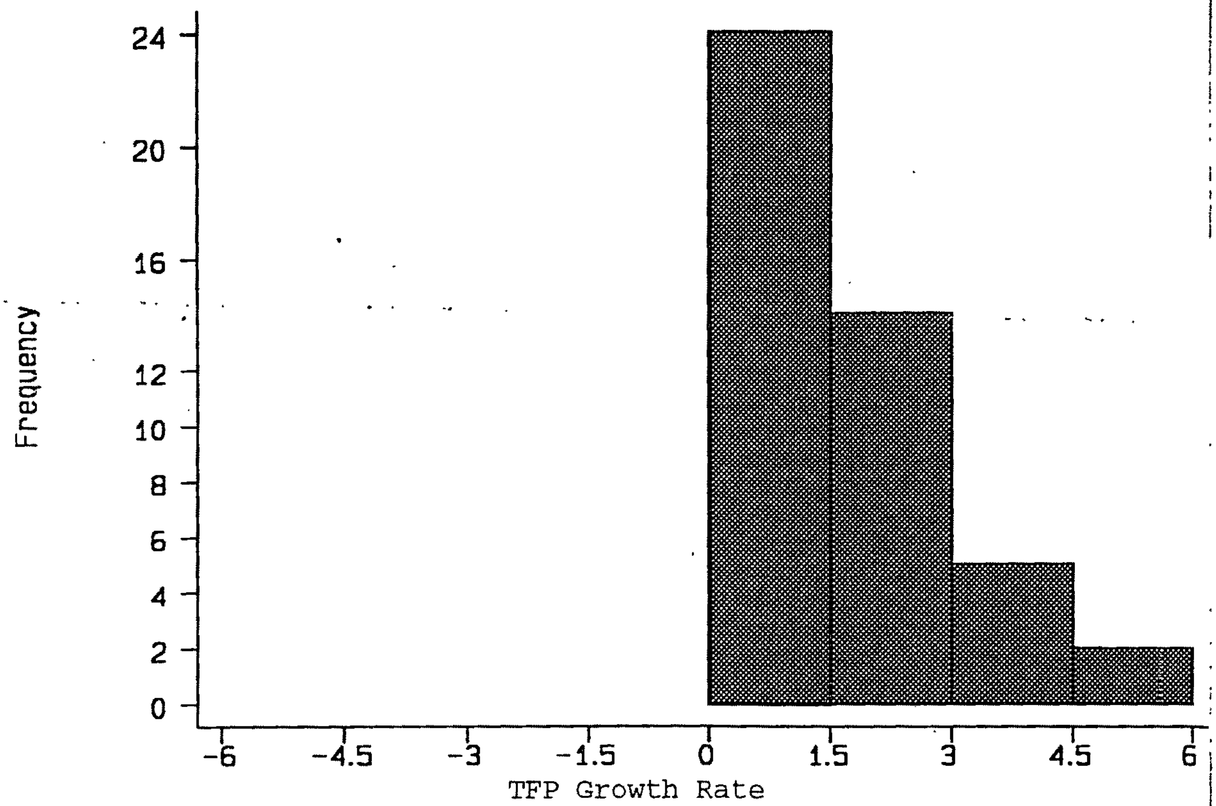


FIGURE 6B. MAXIMUM AVERAGE ANNUAL TFP GROWTH RATE (MAX. ORDINATE OF TFP PROFILE): MEXICAN MANUFACTURING, 1984–1994
FREQUENCY DISTRIBUTION, 44 INDUSTRIAL BRANCHES

value added as his quantity variable, this suggests that, if anything, the substitution of the "true quantity variable" q for observed value added v would have given results with greater dispersion of TFP, and consequently greater overshooting in his sunrise-sunset diagrams.

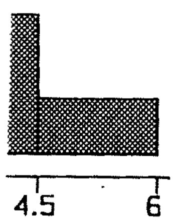
The above demonstration should be taken as merely suggestive. It is not important to me that Torre's results underestimate the variability of the different firms' TFP experience. It is only important that measurement error should not be the principle determinant of those results. On this I feel very confident. In my view, it really is "a jungle out there," with winners and losers in every period—good as well as bad.

As I have noted earlier, we are only just beginning a systematic study of TFP among U.S. firms, so I can offer no display comparable to Torre's.

However, Robles (1997) did examine the experience of 12 firms in the U.S. oil industry. His results are summarized in Figure 7. But Robles tells basically the same story as Torre. Three firms out of the 12 were more than sufficient to generate the real cost reduction experienced by the total group. Half (or almost half) of the firms had negative TFP growth in each period. And the cumulated amount of this negative TFP growth was sizeable when measured against the total TFP performance of the industry.

What I see in TFP performance is quite analogous to what I see in the stock market pages of the newspaper. There are winners and losers every day, every month, and every year. The gains and losses come from all sorts of causes. World price shocks can drive firms into negative TFP performance if the consequent output reductions are greater than the reductions of inputs. So, too, can cyclical or secular declines in demand, including those caused by the successful actions of competitors.

When firms are under stress, they typically fight to stay alive. Maybe they fight for too long in some cases, in the sense that less of society's resources would be wasted if they



'ROFILE):

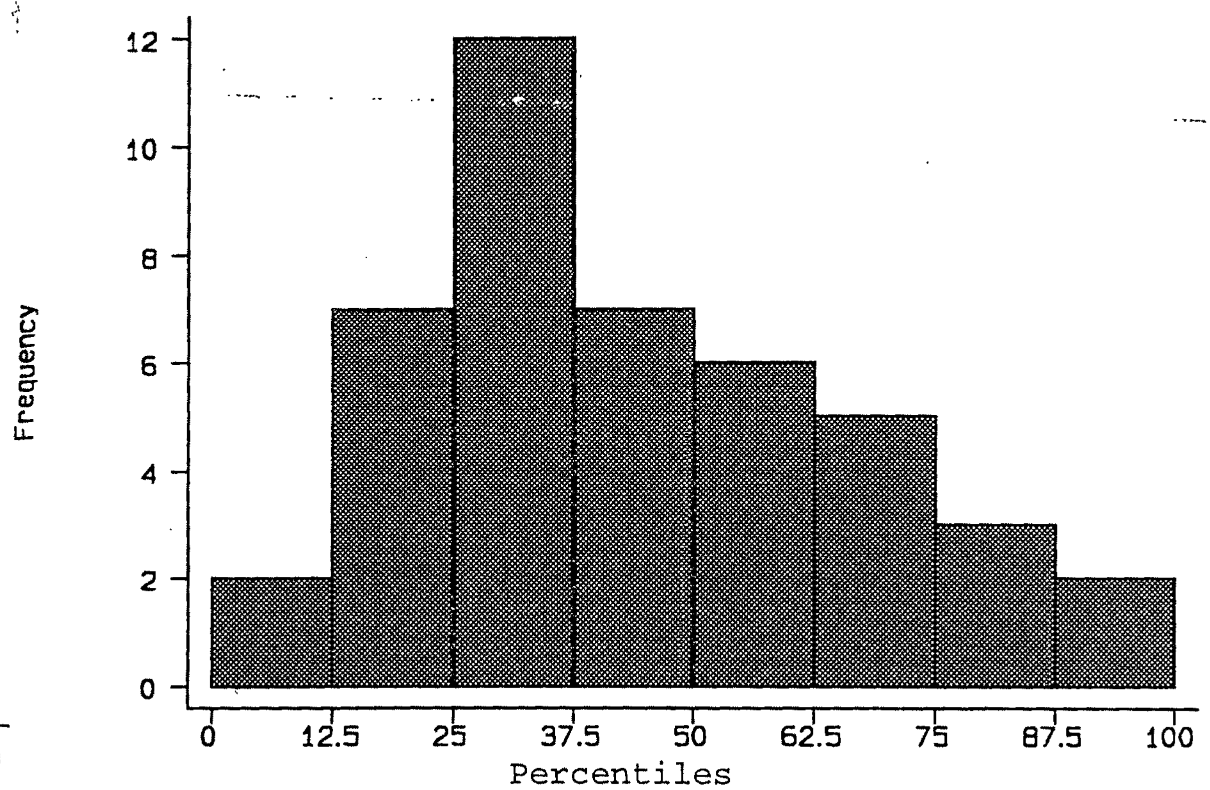


FIGURE 6C. PERCENTAGE WITH POSITIVE TFP GROWTH: MEXICAN MANUFACTURING 1984–1994 FREQUENCY DISTRIBUTION OF PERCENTILES, 44 INDUSTRIAL BRANCHES

story as Tom were to quit earlier in response to a challenge
more than su that turns out to be deadly. But they do not
st reduction e recognize the challenge as deadly, so they
Half (or almo keep struggling to survive. I believe this is part
TFP growth if the nature of entrepreneurs, CEOs, and
d amount of th business leaders in general. They would not be
able when me where they are, doing what they are doing, if
formance of th they were ready to quit at the first sign of a
challenge. They are fighters by nature, and
nce is quite an they probably would not have achieved suc-
market pages cess if they were not.
ners and lose Firms with negative TFP growth may even
every year. The innovators. New challenges come and dif-
l sorts of cause ferent firms think of different ways to respond
rms into negati to them. Some (like Intel and Microsoft) end
quent output up winners; others (Montgomery Ward and
reductions of i Apple?) end up losing. But it may not be that
ecular declines they just waited passively and tried to fight to
d by the succes survive in the face of negative shocks. They
may have had quite innovative ideas, with de-
ss, they typical cent prior probabilities of success, but in the
ney fight for to end success did not come. Thus, negative TFP
ense that less performance can, and I believe often does,
wasted if the come simply from "backing the wrong horse."

To me, Joseph A. Schumpeter's vision (1934) of "creative destruction" captures much of the story. What he is saying is, yes, it's a jungle out there, but the processes of that jungle are at the core of the dynamics of a market-oriented economy. They are what got us to where we are, and they hold the best promise for further progress in the future.[2]

In my opinion, Schumpeter saw through to the essence of the problem, but it is not wise for us to be fatalistic in accepting his vision. We cannot lose by making a major effort to understand the process of TFP improvement where it happens—at the level of the firm. This is all the more true because of the

[2] The idea of creative destruction has come up in recent literature, in a context of formal modeling as distinct from this paper's focus on growth accounting and the intuitive economic interpretation of its results (see Gene M. Grossman and Elhanan Helpman, 1991, 1994; Philippe Aghion and Peter Howitt, 1992). For an econometric study emphasizing the variability of performance among firms, see Jacques Mairesse and Griliches (1993 pp. 200–04).

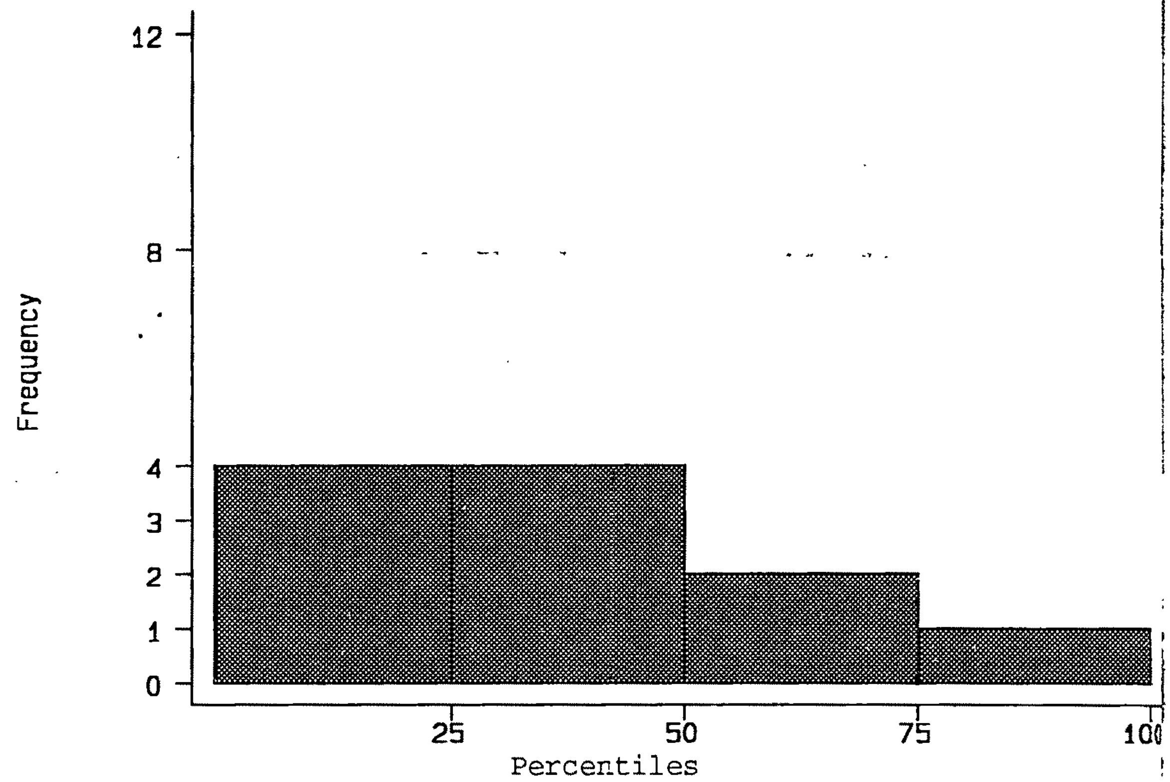


FIGURE 6D. PERCENTILE WHERE CUM. SUM TFP CONTRIBUTION = 100 PERCENT: MEXICAN MANUFACTURING 1984–1994

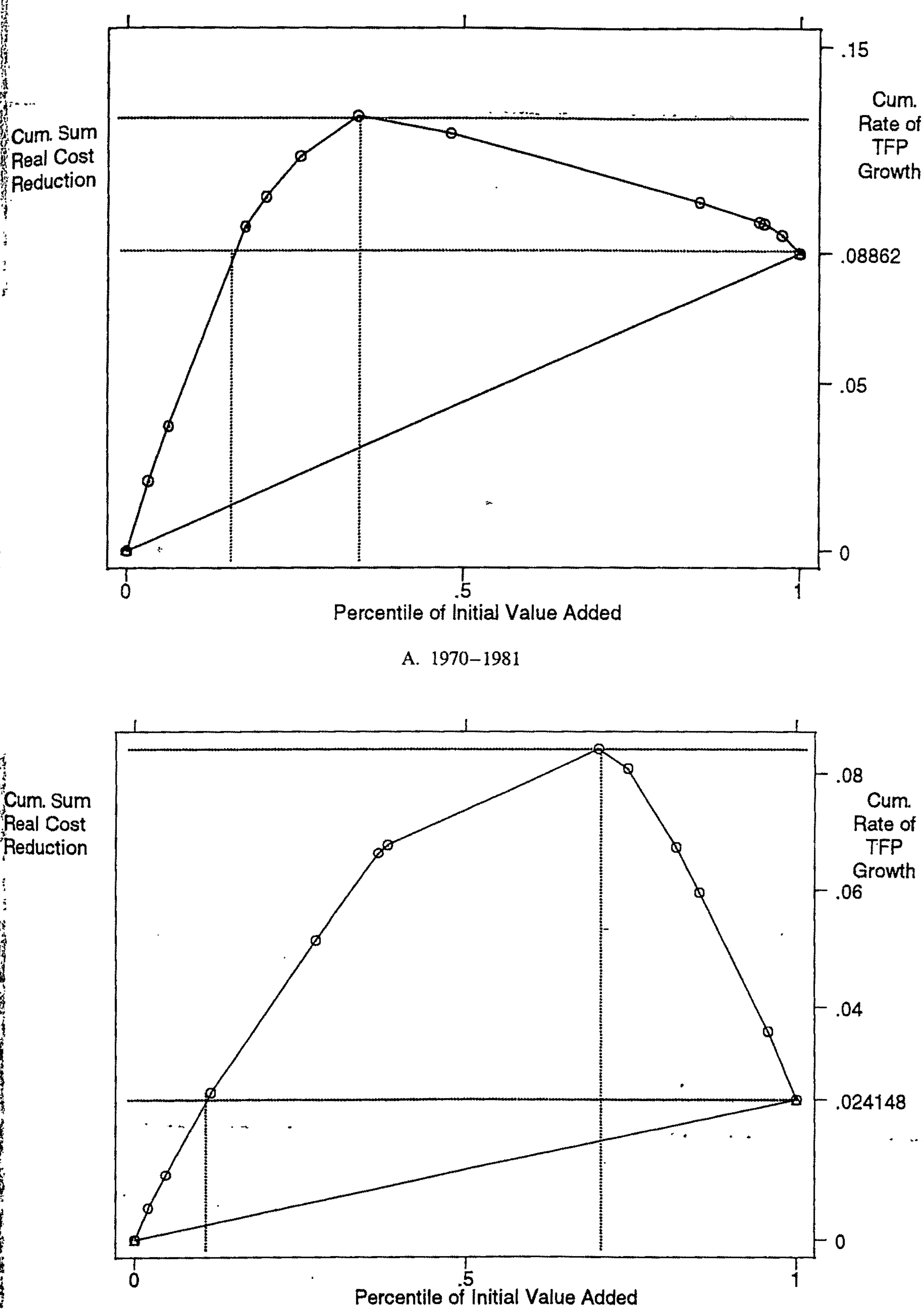
pervasiveness of negative as well as positive TFP performance among the components of almost any aggregate. By learning more about this aspect of the aggregate picture, we may stumble upon ways to "accentuate the positive and eliminate the negative" parts of the TFP story. But that is too quixotic a goal to take as the point of focus right now. To me, the present task is simply to get hold of the huge mass of information that is available at the firm level and squeeze it hard enough to wring out as much understanding and as much insight as we can.

IV. Some Observations on Methods and Research

What I am about to say in this section is not meant to consist of direct implications of what has gone before. Instead, I think of the earlier parts of this paper as building a case for a certain vision of the economy, and of how the forces of growth work within it. This vision in turn leads one to think in different ways not only about the growth process itself but abo how we, as economists, might best advan our study and understanding of it, and ho policies might be molded so as better to pr mote it.

(a) *It is always wise to study the comp nents of growth separately.* The rate of i vestment, the rate of return on capital, the ra of growth of the labor force in numbers or hours worked, the contribution of human ca ital or of the increment in average quality labor, and the residual representing real co reduction—all these are sufficiently differe and potentially sufficiently disjoint from ea other, to merit their being treated separately would give special emphasis to the followi three points.

(i) The worthwhileness of measuring th rate of return and emphasizing its role the growth process.
(ii) The importance of focusing on inves ment rather than saving in studying th



A. 1970–1981

B. 1982–1994

FIGURE 7. TFP GROWTH PROFILES FOR U.S. OIL FIRMS

process of growth. Saving is an interesting topic in its own right, but the more "open economy" is the situation being studied, the less saving has to do with investment. Saving takes on great importance in closed-economy models focused on aggregate growth, in which case it is equal to investment. It gets to be almost meaningless as one focuses on the growth of cities and regions, or on firms and industries.

(iii) The importance of viewing the residual as an umbrella covering real cost reductions of all kinds, and of recognizing that we are closer to home thinking that RCR takes 1001 forms than that it can be well represented by one or two or three aggregate-style variables.

(b) In principle, *the accumulation of human capital by the labor force should be represented in the labor contribution* of the growth equation, or in a bifurcation of this contribution into one due to raw labor, the other to human capital. It is in a term like $\Sigma_i w_i \Delta L_i$ that one captures the shifting skill composition of the labor force. In particular, we capture here the higher wages that are the fruits of investment in education and training, which are the benefits that the workers themselves perceive. These should be kept separate from any externalities education might have.

It is important to try to keep this internalized part of the story out of the residual, so that we can straightforwardly interpret the residual as real cost reduction.

(c) *To study externalities due to education, training, or human capital, we should not be content with broad generalizations* such as "TFP growth is higher in entities with lots of human capital per worker." We should try to figure out how this externality works. Is it higher for *firms* with high incidence of human capital? Is it higher for industries or sectors? Or are human capital externalities more spatial in nature, making more efficient the economic life of the cities, provinces, states, or nations which have high concentrations of human capital? And if this is a fruitful trail to pursue, at what type and size of geographical units do these externalities typically work?

(d) The same goes for economies of scale. *We should not be satisfied with vague attributions of economies of scale, say, at the level of the national economy.* Instead, we should pursue the matter. If the economies of scale are national, through what channels do they work, and what evidence do we have to look at to see them in operation? In particular, what is their connection to real cost reductions where they really happen—i.e., at the level of the firm? Economies of scale at the levels of the firm and the industry are easier to visualize. Here, too, however, the task is to check them out—to see if the real cost reductions of firms are linked to the initial sizes of those firms themselves, or of the industries in which they operate, and of the direction (up or down) in which output is moving.

(e) Perhaps most important of all, we should really try to *take full advantage of evidence at the firm level.* I think particularly of identifying considerable numbers of outstanding cases of TFP improvements and TFP decline, and studying them one by one to try to ferret out the sources of their big real cost reductions and real cost increases. You can be pretty sure, if there have been big real cost reductions in a firm, some people in that firm have a pretty good idea of where those reductions came from and how they were accomplished. By capturing this grassroots evidence, we can put some added discipline into our ruminations about the nature of TFP at the aggregate level. We must follow up on the sort of work pioneered by Jacob Schmookler (1966) and Edwin Mansfield (1995). In general, our aggregate story should be compatible with, and comfortably contain, what we see at the grassroots level. In particular our overall picture of TFP improvement should comfortably accept the overwhelming evidence of the "mushrooms" rather than "yeast" nature of the process.

(f) *Special urgency applies to the study of declining total factor productivity at both the firm and the industry levels.* The pervasiveness of declining TFP is perhaps the most profound conclusion to emerge from the empirical links that I have reported here. As a profession, we obviously have been aware of its existence at the industry level for virtually all studies that

give a breakdown by industry reveal it. Yet to my knowledge, we have barely scratched the surface in studying it. I find it hard to think of more fertile soil for future research on the process of economic growth.

(g) I do not think that we gain much by trying to express the relation between policies and economic growth by a series of regressions. *Cross-country growth regressions seem hopelessly naive* to longtime observers of the growth process like myself. To us, there is too much to question in regression lines that draw much of their slope from the differences between Sudan and Switzerland, between Bangladesh and Brazil, or between Ceylon and Canada. In contrast, it seems much more sensible to look at episodes within individual countries and to search for common elements that characterize the passage from bad to good growth experiences within each of the number of countries, and for those elements that seem to describe the good growth experiences on the one hand and the bad ones on the other. I think we can reach in this way a good appreciation of the nature of the growth process, without resorting to the straitjacket of regression lines that seem to draw from comparisons among very disparate countries, lessons that are supposed to be meaningful for countries like Bangladesh, Ceylon, and the Sudan—as well as others at different levels—as each strives to take the next upward step in the climb toward modernization.

My view of cross-country growth regressions is somewhat less negative to the extent that they focus on the components of growth (rate of investment, rate of return, and real cost reduction in particular) rather than on the overall growth rate. There is also a subtle distinction to be drawn between two ways of presenting cross-country regressions—(i) as "explaining" why and how some countries grow faster than others (not recommended), and (ii) as simply summarizing a series of "stylized facts" describing the experience we observe (far preferable, and not just for its being more modest in its claims).

V. Some Policy Implications

In approaching the question of the influence of policy on real cost reduction in particular, and, to a degree even on economic growth in general, I believe that the key words are "obstructing" and "enabling." We know from sad experience how easy it is for governments to adopt policies that get in the way of economic growth and even turn it negative. We know, too, that there is no "silver bullet," no single magic key that by itself opens the door to a paradise of prosperity and growth. Broadly speaking, the easiest starting point for a successful growth experience is a once-prosperous economy that has suffered from bad policies. Releasing that economy from its trammels, correcting an accumulation of past mistakes, can sometimes set in motion a prolonged episode of astounding growth. A shift from policies that obstruct to policies that enable growth seems to lie at the heart of growth "miracles" like those of Taiwan, Spain, Korea, Brazil, Indonesia, Malaysia, and China (among others).

The springboard for the following listing of policy implications is the interpretation of the growth residual as representing real cost reduction and the ready acceptance that in the real world RCR comes in 1001 different forms.

(a) The first key observation is that *people must perceive real costs in order to reduce them.* Hence, policies that impede the accurate perception of real costs are ipso facto inimical to growth. *Inflation* is the most obvious, probably the most pervasive, and almost *certainly the most noxious of such policies.* If I have any expertise based on experience in economics, it has to be in the first-hand observation of processes of serious inflation. So I ask you to take my word for it: the most serious cost of inflation is *not* a triangle or a trapezoid under the demand curve for real cash balances, nor is it the inflation tax. The most serious cost of inflation *is the blurring of economic agents' perceptions of relative prices.* This happens because individual prices adjust in different ways and at very different rates. A high product price and a low input cost normally is an invitation clamoring for new investments to be made. This is not so during a serious inflation, when such a signal can easily turn out to be "here today, gone next month" as both product and input prices continue on their separate paths of adjustment to

P-11029

the ongoing inflation. Without exception, in my own observations, the higher the rate of inflation, the worse is its effect in blurring agents' perceptions of relative prices. In an inflation at, say, 20 percent to 50 percent per year, people see prices as in a morning haze; in one of 20 percent to 50 percent per month, they see them as in a London fog. Many empirical studies exist showing that serious inflations are seriously inimical to growth. (See William Easterly, 1996.) The clouding of perceptions of relative prices is an important reason why—for it gets in the way of successful real cost reductions at the level of the individual firm.

Inflation also inhibits growth in other, perhaps more obvious ways:

(i) by diverting energies from more productive activities to the search for mechanisms of inflation protection;
(ii) by reducing (often very drastically) people's real monetary balances, thus impacting negatively on the real amount of credit the banking system provides to the productive sector; and
(iii) somewhat related to both (a) and (b), by causing people (both "here" and abroad) to invest abroad some of the funds they would otherwise have invested "here," or (what is very close to the same thing) by accumulating hoards of hard currencies as an inflation hedge.

(b) A second policy implication is, in the words of my friend and longtime collaborator Ernesto Fontaine, *avoid "prices that lie"* (*precios mentirosos*). Talking about inflation, we focus on the blurring of the signals that the price system gives; here we focus on its giving wrong signals due to distortions that have been introduced, usually as a direct consequence of government policies. No good can or did come, in terms of economic efficiency, from tariffs of 50 percent and 100 percent and more, giving effective protection often of 200 percent and 300 percent and more. Nor can growth be fostered by heavy-handed price controls and interventions in credit markets.

I am not being a religious purist here—just as big distortions have big costs, small distortions typically have small costs, and all economies are distorted to some degree. The message here is *that economies have to pay the price for the level of distortions they choose to have,* and that one of the important components of that price is that *distortions create situations where what is truly a saving of private costs is not a genuine saving of costs from the point of view of the economy as a whole.*

(c) Just as bad, and often even worse than direct distortions, are the excess costs imposed on an economy by *ill-conceived regulations and bureaucratic hurdles.* Hernando DeSoto (1989) has made the exposure of these trammels in Peru into what has become virtually his life's work. Clear rules of the game are an essential and integral part of a well-functioning market economy, but all too easily these get supplemented by others that make investment, production, marketing, sales, new product development, etc., more costly. Labor laws have been particularly troublesome, often adding artificially to the cost of labor and giving firms a strong incentive to avoid hiring new workers, simply because of the high costs associated with any later dismissal of them.

But there are loads of other items—*the need for approvals*, sometimes a dozen or more before undertaking some investment or some new venture; regulations that one way or another impede new entry, so as to protect strong vested interests (small retailers being protected against supermarkets in many countries); and the complexity of tax codes and their enforcement, which imposes large compliance costs on business firms and individuals. Somehow, countries interested in promoting growth should find ways of paring their regulatory frameworks down to those rules and requirements that are really justifiable in terms of their costs and benefits to the economy and society at large.

(d) Although *international trade distortions* (tariffs, quotas, licenses, prohibitions, etc.) might be subsumed under points (b) and (c), their importance merits a separate heading. The move to openness (from a protectionism that sometimes bordered on autarchy) has been one of the main hallmarks of the growth miracles of the past half-century [se

Sebastian Edwards (1993) and Anne O. Krueger (1985, 1997)]. Just as inflation has costs beyond the area under the demand curve for real cash balances, so protectionism seems to have burdens that go beyond the standard triangle-trapezoid-rectangle measures of the costs of trade distortions. There are at least two quite natural explanations: first, that openness helps grease the wheels of the international transfer of more modern technologies, and second, that firms that may once have relaxed in ease and comfort behind high protective barriers end up having to sink or swim once forced to compete in a much more open-economy setting. Under either explanation, trade liberalization opens up new paths of real cost reduction, thus providing additional impetus to economic growth.

(e) The recent wave of *privatizations* among both developed and developing economies may have important effects in enabling real cost reductions that otherwise might have been delayed, or not have happened at all. It is, I believe, fair to say that in most countries state-owned enterprises operate under a series of constraints that seriously get in the way of real cost minimization in a comparative-static sense and of real cost reduction in a dynamic sense. These constraints sometimes limit the salaries of executives, sometimes impose onerous conditions on the firm as it employs lower-skilled workers, often limit the capacity of the firm to shut down inefficient lines of production, and almost always make it difficult to fire workers, etc. To my mind, however, perhaps the worst attribute of state-owned enterprises is the ethos that often evolves inside of them—an ethos where middle managers are well advised to "leave well enough alone," "not rock the boat," and "not invite trouble." This ethos flies in the face of a vision of the growth process that gives a huge role to the search for real cost reductions at the grass-roots level, and that recognizes the tumult that accompanies "creative destruction" in all its forms. I thus must applaud the contemporary trend toward privatization. If I harbor any qualms in this connection, they concern the degree to which many privatizations have been carried out in too much haste and with too little care, often motivated by purely fiscal considerations rather than by a general search for economic efficiency. This may have led to gratuitous transfers of wealth in some instances and to the planting of newly private enterprises in soil that was not properly prepared (e.g., still lacking a sound regulatory framework for electricity rates, or intelligent rules promoting competitiveness in at least some aspects of telecommunications, etc.)

(f) One cannot complete a list like this without mentioning something that most of us simply take for granted—*a sound legal and institutional framework in which individuals are protected against arbitrary incursions on their property and other economic rights*. This very basic point—recently much emphasized by Douglass C. North (1990), Robert J. Barrow and Xavier Sala-i-Martin (1994), Mancur Olson, Jr. (1996), and Barro (1997)—is at least potentially a vital element for a sustained process of successful economic growth. If it is true that spurts of growth have sometimes occurred in the absence of such a framework, it is also true that most cases of sustained growth over long periods of time have benefitted from a sound institutional and legal environment.

(g) Somewhat related to the above is the element of *political consensus concerning the broad outlines of economic policy*. We have learned from experience that very admirable policy reforms can take place, yet end up having little effect. This can happen because a new government comes in and reverses the reform. But it can also happen because people fear that a new government will come in and reverse the reforms later on. At the moment, the Chilean economy is one of the jewels of economic growth (and general economic success) in Latin America. Many people point to the thoroughness and pervasiveness of Chile's economic reforms over the last two decades or so. But not so many point to the fact that the reform package has remained essentially intact through several changes of ministers, and even more important, through two presidential elections in which the winners came from the opposite side of the political fence from the government that initiated the reforms. The confidence in the economic order of things instilled by this sequential endorsement of the basic framework of economic policy has to be

one of the important reasons for Chile's continued, very impressive economic performance. And it is important, also, in the context of this paper. Living in a world in which real cost reductions are a key dynamic force producing economic growth, we must look to the motivations and preoccupations of those who take the critical decisions at the level of the firm. For these decisions, it is not only important that the policy framework be good now; the expectation that it will stay good in the future is also important. Otherwise, investments will tend to be limited to those with short horizons and payment periods, and much soil, fertile with longer-term economic opportunities, will go unplowed.

VI. A Vision of the Growth Process

Let me now try to summarize my own vision of the growth process—the major elements of which have been presented here. In the first place we have the five standard pillars of growth—the rate of increase in the labor force, the rate of increase in the stock of human capital, the increase in the capital stock (net investment as a fraction of value added), the rate of return which that investment will yield (or can be expected to yield) and, last but not least, real cost reductions stemming from 1001 different sources.

Commenting on these in turn, I would note that increases in the labor force have taken on new meaning in many countries as labor force participation rates (particularly those of women) have increased. Whereas with a constant participation rate, the growth rate of the labor force is just a proxy for the growth rate of population, important increases in labor force participation can lead, just by themselves, to significant increases in measured real income per head.

Concerning increases in the stock of human capital, my conviction is that most of their contribution to growth is, on the whole, well measured by market wages, as in the expression $\Sigma_i w_i \Delta L_i$. This does not deny the existence of externalities due to an increased human capital stock; it simply judges their influence on the growth rate to be modest in comparison with the effects of education, training, learning-by-doing, etc., that can be (and typically are) internalized by those who receive them. We therefore look for the effects of human capital accumulation mainly in the term $\Sigma_i w_i \Delta L_i$, and only (via externalities) as one of many elements underlying the growth residual R'.

The rate of investment is a veteran on the stage of growth analysis. What I would emphasize here is the importance of maintaining a clear separation between the rate of investment and the rate of saving. Models (like those of the representative consumer) in which saving and investment are always equal are not much use even for analysis at the national level in our modern, interdependent world. They are even less useful as one goes down to smaller geographical regions, and simply cease to make sense as one studies the growth process at the level of the industry and the firm.

The rate of return to investment has in many ways been the orphan of our growth analysis, having been masked from view by our typical representation of capital's contribution to the growth rate as $s_k(\Delta K/K)$. Here the rate of return $(\rho + \delta)$ is totally hidden from view. I deeply urge that more of us get into the habit of representing this same term as $(\rho + \delta)(\Delta K/y)$. I want to see more attention paid to the rate of return because it plays such a central role in the motivation of economic agents, and also because changes in it are such an important element in understanding and explaining the growth residual, R'. Table 3 helps explain why I feel this way. This table is adapted from Harald Beyer's (1996) work. He carried out an analysis of the growth experiences of 32 countries ranging from Sri Lanka to the United States on the income scale, and from Iceland to Australia on the geographic scale. In Table 3 we present results for his ten countries with the highest and for his ten countries with the lowest GDP growth rates from 1971–1991. In the second column the calculated average annual rate of return is shown. In the third column we have capital's contribution to the growth rate $[= (\rho + \delta)(\Delta K/y)]$, and in the final column the estimated average annual rate of TFP growth, all over the same time period.

Table 3 shows an unequivocal tendency for fast-growing countries to be experiencing high rates of return as well as high capital contributions and high rates of TFP improvement

TABLE 3—GROWTH RATES, RATES OF RETURN, AND RATES OF TFP IMPROVEMENT (SELECTED FROM A SAMPLE OF 32 COUNTRIES, 1971–1991)

Ten fastest-growing countries	GDP growth rate	Rate of return	Capital cont. to growth rate	TFP growth rate
Taiwan	8.83	15.0	3.81	3.68
Korea	8.47	13.2	4.30	2.38
Thailand	7.65	12.5	3.68	2.96
Hong Kong	7.91	20.0	3.56	2.28
Ecuador	5.58	14.0	2.70	0.36
Cyprus	5.12	10.6	2.99	1.92
Zimbabwe	4.62	13.6	2.42	0.97
Colombia	4.43	11.3	1.99	0.74
Iceland	4.35	9.4	1.95	1.77
Ireland	4.12	6.7	1.70	0.36
Median	5.35	12.85	2.84	1.83
Mean	6.10	12.63	2.91	1.74
Ten slowest-growing countries	**GDP growth rate**	**Rate of return**	**Capital cont. to growth rate**	**TFP growth rate**
Austria	2.87	5.1	1.13	1.29
France	2.80	6.1	1.21	0.99
Germany	2.60	6.3	0.97	1.29
Belgium	2.56	6.8	1.06	1.60
Netherlands	2.52	7.0	1.12	0.83
United States	2.52	9.1	1.20	0.23
South Africa	2.16	7.5	1.58	−0.97
Denmark	2.15	7.5	1.01	0.82
United Kingdom	2.12	9.6	0.95	0.22
Sweden	1.84	4.3	0.66	0.24
Median	2.52	6.90	1.09	0.825
Mean	2.41	6.93	1.09	0.661

Source: Beyer (1996), Tables III.1.1 through III.1.32; also Appendix I for rates of return.

…is is all the more interesting because in the …culation of TFP a higher *level* of the rate of …urn operates to reduce the calculated TFP …e., $\Delta\rho$ is a positive component of R' and …uld presumably be positively correlated …h it,[3] but R' is found by subtracting $\rho\Delta K$ …m Δy; hence, in a sense, the *level of* ρ should presumably be negatively correlated with R'). What we are seeing here, in my opinion, is a genuine syndrome in which all sorts of good things go together. Strong real

[3] The standard expression for the residual, $R = \bar{p}dy -$ … $- (\bar{\rho} + \delta)dK$ has a "dual" representation, which is: $R = \bar{L}dw + \bar{K}d(\rho + \delta) - \bar{y}dp$. This form simply says that the fruits of real cost reduction have to go somewhere—either to workers ($\bar{L}dw$) or to owners of capital $[\bar{K}d(\rho + \delta)]$ or, in the form of lower prices to the activity's customers ($-\bar{y}dp$).

cost reductions and high rates of return create attractive investment opportunities which, when acted upon, bring about a high capital contribution to growth. It should be no surprise that under such circumstances the GDP growth rate itself tends to be high. It should likewise be no surprise that the opposite syndrome—with weak real cost reductions and low rates of return producing fewer interesting investment opportunities—should end up being associated with a low capital contribution and a low GDP growth rate.

Finally, we come to the residual R' itself. To me, the biggest message here is to recognize the *multiplicity of sources* from which it can (and I believe does) come, and the *additivity* that nevertheless is its attribute. I think that the term real cost reduction very neatly captures both these aspects in a way that renders it preferable to terms like TFP improvement and technical advance—preferable not in the sense of a mechanical definition (for which all three are equally good), but in the sense of better conveying the underlying nature of the process to one's listeners or readers.

The next step is to recognize that of the five main pillars, at least three (the rate of investment, the rate of return, and real cost reduction) are key foci of decision-making processes at the level of the firm. I cannot escape the conclusion that the great bulk of the action associated with the growth process takes place at the level of the firm. Hence, I feel we should focus much more study than we have in the past on what happens at this level. And when we are not working at the firm level, we should pay a lot of attention to what happens at lesser levels of disaggregation like industries and industrial branches.

Key insights flow from taking this kind of focus. Few economists are aware of the pervasiveness with which sunrise-sunset diagrams are characterized by overshooting, or of the importance that firms or industries with real cost increases (i.e., reductions in TFP) play in determining the aggregate rates of real cost reduction that we see in such diagrams. Here we have only scratched the surface in digesting the evidence. But I find impressive the degree to which the data of Table 3 seem to point to a growth syndrome in which high rates of return, high rates of investment, high rates of real cost reduction, and high rates of output growth all go together. I see in this result the likelihood that real cost reductions are the big driving force, generating high rates of return and calling forth high rates of investment and high output growth. This interpretation is compatible with many exercises that I have performed over the years in which I have tried to contrast high-growth with low-growth experiences. In such exercises, as in Table 3, the difference in rates of real cost reduction has typically been a major "source" of the difference in growth rates.

Also impressive in the analysis of the Jorgenson data is the degree to which the varying experiences of U.S. manufacturing in different decades derives from different degrees of bad experience (real cost increases) rather than different degrees of good experience (real cost reduction). It is as if the creative part of Schumpeter's "creative destruction" was more steady (for these decades in U.S. manufacturing) than the destructive part, whetting our (or at least my) appetite to look deeper in inquiring into why this was so.

The Mexican data at the firm level were somewhat more recalcitrant than the Jorgenson data by industrial branch, but they nonetheless give us a clear picture of lots of winners and lots of losers, with the losers being strongly characterized by falling real value added and or by falling real rates of return.[4]

[4] Of Torre's observations with negative TFP growth over three-fourths had negative growth in real value added, and over one-half had falling real rates of return. Less than 15 percent showed increases both in real value added *and* in their real rate of return. Many of Torre's "anomalies" of negative TFP improvement together with positive real output growth stem from very high rates of return ($\bar{\rho}$) being imputed to the observed increases (ΔK) in the capital stock.

This points to a problem that extends to all (or nearly all) growth-accounting frameworks. Implicitly, they assign to new investment a marginal product based on the average observed gross rate of return ($\bar{\rho} + \delta$) or average observed capital share s_k. This makes little sense in cases where the observed $\bar{\rho}$ is far above or far below the going market rates. Firms earning 50 percent real return in their historical investments are very unlikely to be "requiring" such a high expected yield on new investments. Similarly, firms going through periods of actual accounting losses may often still be investing, but it is absurd to think they are expecting (or typically getting) negative returns on their new investments (ΔK). There are, I think, good reasons for us to experiment with alternative ways of sel-

The role of policy in this vision of the growth process is an "enabling" one. By creating the circumstances where firms can quickly and accurately predict opportunities for real cost reduction and act on them, governments can guide the economy toward an enhanced contribution of RCR to growth. By rationalizing and/or eliminating barriers and controls, they may also lead to an increased pace of investment and increased rates of return. In this view, the connection between good policy and growth is not mechanical, and thus cannot easily be captured in regression-type analysis, but that does not stop it from being of vital importance.

I want to give special weight to another role of "growth-enabling" policy actions, which has to do with how policies, the effects of which might typically be considered to be comparative static, can nonetheless turn out to influence economic growth rates over extended periods of time.

This story begins with a recognition that most developing countries have typically used production techniques that were "backward" in relation to those used by the advanced economies. One way to verify this is to imagine integrally replicating, say, a U.S. factory in India, and manning it there with Indian workers equivalent in skill to their U.S. counterparts. The combination of lower construction costs and much lower operating costs (mainly wages) would permit this hypothetical new Indian firm to undercut the prices of *both* the U.S. firm of which it was a copy *and* the typical Indian firm currently active in the same industry. This says that the typical Indian firm is operating on an "inferior" production function. If, as I believe, the difference in efficiency between U.S. and developing-country firms is typically large, there is much room for quite rapid improvements in the developing countries as they learn how to "adopt and adapt" already-known techniques from the advanced countries.

I would assume that the incentives for "convergence" are always present, but that they have typically run into barriers and trammels of many kinds in the poor countries of the world. Reducing the barriers and loosening the trammels permits the more rapid convergence to techniques that are closer to the frontier of knowledge.

The way I see the influence of policy in growth, it is simply not true that implementing enabling policies typically permits a quantum jump from the old to the truly modern. It is more accurate to describe it as speeding up what will in any case be a very lengthy process. Personally, I like the analogy to a hydraulic system in which a large vessel with a high water level and lots of water is connected to a much smaller and narrower vessel with a much lower level of water. Physical laws dictate a tendency for the water levels to equalize in the end. But this can take a very long time if the tube connecting the two vessels is tiny, or is clogged by extraneous matter.

The policies that we consider good for growth have the attribute, in this analogy, of removing the extraneous matter and/or enlarging the connecting tube. But even with the best modernizing policies, the tube remains small enough so that it takes many decades for a country to pass from poor to middle income or from middle income to rich. If somehow the hydraulic connection could be made so large as to bring about an almost instantaneous full adjustment of the water level, then we would say that good policies mainly represent level adjustments. But observing even the best of real-world growth experiences, I think we have to conclude that the adjustment is going to be extended over a lengthy period in any event, thus causing the big observable result of better policies to be a higher growth rate over an extended period rather than a discrete jump to a totally different level.

This way of looking at the world also leads to some observations on the current literature dealing with convergence. I have long been mystified by allusions to an ultimate convergence of growth rates among countries, or an ultimate convergence of levels of output per head. To me, the natural convergence is product by product, not country by country. And

... the rate of return (ρ) imputed to new investments. ... uments can be developed for using arbitrary but "sensible" rates (like 10, 15, or 20 percent), or for applying ... investments by individual firms and/or industries rates of return $\bar{\rho}$ that are obtained not from those specific firms or industries but from broad sectoral (e.g., manufacturing) or economywide averages. I am pursuing these avenues in work that will be reported at a later time.

among products there may be some where current techniques will never be further improved. Those products will have no real cost reductions or TFP improvements in the future, while others will enjoy huge advances of productivity. My guess is that unlucky countries (Bhutan, Nepal, Mongolia?) may always lag way behind the pack, while luckier ones (Taiwan, Argentina, Brazil?) may one day hope to be among the world's leaders. So convergence comes through as a general tendency, and quite surely a general possibility, for the production techniques used in making any given product to improve as enterprises using "backward" techniques learn of better ones, and even more important, learn about how to put them into practice. I do not believe that much more than this can be said of convergence as a real-world force. Wage rates for given types of labor tend toward a rough equality across regions in a country because of the ease of migration in response to perceived wage differences. The forces at work internationally are both weaker and more complex, but the big message here is that the improvement of technique in any one industry does nothing to improve the wages of labor in just that industry. After an improvement, the industry may end up hiring more or less labor, but will presumably choose the amount of labor so as to bring marginal productivity into correspondence with market wages. So technological improvement has an effect on wages via supply and demand in the national labor market, not through any direct link from technical improvement to wages (for given skills, etc.) at the level of the industry or the firm.

As my final point, let me return to the thought that the justification for perfecting the functioning of the market system does not lie only in reducing the efficiency costs associated with each period's operation of the economy. Perfecting a country's economic policy does not only cause it to move from a path at around, say, 90 percent of its potential output to another equal to 95 percent of potential, with the time path of potential output being somehow given in advance. That gain would certainly be a worthwhile gain, and it would amply justify a lot of hard work involved in achieving it. But that gain is still fundamentally comparative static in nature.

What I hope to have evoked in this paper is a sense that the perfecting of economic processes can also in nearly all cases be justified as greasing the wheels of the constant search for new avenues of real cost reduction. To the extent that economic reforms do so, they become vehicles for bringing an economy to a point where, year after year, new, cheaper, and better ways are found of doing things, not just in so-called "production" but also in such mundane areas as merchandising, sales, finance, insurance, and many more.

Some years ago, in a book that I edited called *World Economic Growth* (1984), I wrote an essay called "Economic Policy and Economic Growth," in which I listed "13 lessons" that I thought followed from the papers presented in the volume, recounting the growth experiences of countries as disparate as Ghana and Taiwan, or Japan and Sweden. These lessons basically focused on thinking about policies in terms of their economic costs and benefits—could easily be read as a reprise of the old comparative-static story. But they were not meant as such—it is quite relevant that they appeared in an essay concerned with world economic growth. The point was that these sensible policies emerged as part of a consensus of serious economists, each an expert in his particular country's history, focusing attention on the process of economic growth.

A few years later, and as the theme of a different concept, John Williamson (1990) coined the term, the "Washington consensus." Williamson listed ten points, covering territory very similar to my 13 lessons. He also produced a pithy summary that captures the essential thrust of both his and my listings: "Macroeconomic prudence, outward orientation, and domestic liberalization." He, too, and the members of the Washington professional establishment whose apparent consensus led to Williamson's list, were not just thinking of comparative-static gains as they reached their conclusions about policy. They were thinking about ways to move economies from slow growth, stagnation, and even in some cases negative growth, to a healthy, prosperous flowering of economic progress. Similar views were more recently affirmed by Stanley Fischer (1993).

To me, the dynamics of real cost reduction are at the very least an important piece of what

people have in mind when they list efficiency-oriented policies as essential ingredients of a program promoting economic growth. It is policies of this type that give the right signals to the CEOs and the managers down the line, that take away trammels that impede their quest for real cost reductions, and that create an environment in which Schumpeter's process of "creative destruction" can work its wonders.

APPENDIX ON METHODOLOGY

The vision of the growth process presented in this paper leads one almost inevitably to some methodological twists—twists which, if they are not new, at least differ in significant ways from what I take to be the most common practice in breaking economic growth down into components.

(a) To measure the real rate of return to capital, one must express the numerator (real dollars of return) and the denominator (the capital stock) in the same units. The most efficient way to do this is to measure both output (value added) and the capital stock in units of the GDP deflator. That way one is sure that the outputs of all the subaggregates in the economy add up to the GDP, and one also satisfies the requirement that the capital and return be measured in the same units. This is also the way cash flows would be deflated in a standard, *ex post* project evaluation. When this is done at the aggregate level the contribution of capital to growth is $(\rho + \delta)\Delta K$ where ρ is an aggregate rate of return to capital, δ the depreciation rate (including obsolescence), and ΔK the net increment to the capital stock in the period in question. At a subaggregate level, the contribution of capital to growth in activity j is $(\rho_j + \delta_j)\Delta K_j$. At both levels, we find that high rates of return are an important component of most successful growth episodes.

(b) To capture the great diversity of the labor factor, we would like to have a very fine breakdown of labor into categories (indexed by i). The labor contribution is then $\Sigma_i w_i \Delta L_i$, where w_i represents the real wage of category i and ΔL_i the change in hours worked by category i. Since the number of relevant categories of labor is huge, any such breakdown is difficult, and gets more difficult as one disaggregates from economy to sector to branch to industry and to firm. This Gordian knot can be cut by a simple assumption, similar to what is done in most countries to convert residential construction to real terms. There, one builds a price index of a "standard house" p_h^*, and then obtains a quantum of construction C^* by dividing total construction outlays by the price of the standard house. In the resulting aggregate, each individual residence (i) gets attributed a quantum of housing equal to p_i/p_h^*. In this work I define a standard wage w^*, which I assign to "standard labor" or "raw labor." The excess of anybody's actual wage over w^* is attributed to human capital. The returns to natural ability, as well as to formal education, training, and experience belong there, under this interpretation. High returns due to a distorted wage structure are not appropriately attributable to human capital, but the methodology would nonetheless be correct in attributing to the affected labor a marginal productivity that is measured by the distorted high wage.

The "labor contribution" as measured by $w^*\Delta L^*$ is equal to $w^*(\Sigma_i (w_i/w^*)\Delta L_i + \Sigma_i L_i \Delta(w_i/w^*))$. The second term will be zero if the structure of relative wages remains constant, or even if the weighted average premium does not change. Any changes in the weighted average premium will cause the calculated residual to be different from those calculated by other methods.

The "two-deflator method" is characterized by the use of a single numeraire-deflator (say, the GDP deflator), by the treatment of the quantum of output as value added divided by the numeraire-deflator, and by the use of a standard wage w^* and a quantum of labor L^* equal to the wages bill divided by w^*. This is the method used by Beyer (1996), Robles (1997), and Torre (1997) in their work reported in this paper.

It goes without saying that the two-deflator method is rough. But it is also tremendously robust and easily applied. I think of it as being really designed for use at the firm level, where very commonly we can get data in value added, on gross investment, and on the wages bill, but know nothing (from standard sources) about the quantum of output or about the number of total hours worked (or even the total number of employees used). This opening of wide new vistas, of huge new data sets, is what

I consider the strongest argument for the two-deflator approach.

But there are other pluses as well. First, at the aggregate economy level, the two-deflator approach comes very close to the traditional approach. In rate-of-growth terms, we have $(R^*/y) = g_y - s_k g_{k^*} - s_L g_{L^*}$, compared with $(R/y) = g_y - s_k g_k - s_L g_N$, where g refers to growth rate, and y to GDP. K^* differs from K in being built up from gross investment deflated by the GDP deflator, while K is built up from gross investment deflated by the investment deflator of GDP. L^* is in principle much more refined than N (number of workers), but its measurement can be influenced by a widening or narrowing skill premium. (R^*/y) likewise differs from Jorgenson's residual $(R'/y) = g_y - \Sigma_j s_{kj} g_{kj} - \Sigma_i s_{\ell} g_{\ell i}$ mainly in his use of different capital deflators for different categories of capital. (The implicit labor breakdown in L^* is much finer even than Jorgenson's, but Jorgenson does not have to contend with the possibilities of widening or narrowing skill premiums—at least not among the labor categories he works with, which are typically broken down by gender, age, education, occupation, industry, and employment status.)

The bottom line is that when Beyer compares his two-deflator results, at the national aggregate level, with those of others using different methods, he finds on the whole only modest differences. [See Beyer (1996); Harberger (1998).]

When one uses the two-deflator method at the industry level, one often has the possibility to adjust the quantity variable so as to make it correspond with that of the more traditional approach. Thus, we may start by using dy_j^* ($= p_j dy_j + y_j dp_j$) as the quantity variable, and calculate a residual R_j^* using that concept. Then we may get an adjusted residual by taking $R_j^* - y_j dp_j$. This is easy to do so long as we have decent data on dp_j, the relative price of j, which we often do at the branch or industry level. When Jorgenson's residuals are compared with the two-deflator residuals, with and without price adjustment, I find the differences without adjustment to be small enough to be quite acceptable. With adjustment, the degree of agreement between the two approaches is quite notable (85 percent of differences less than one percentage point of per annum growth). [See Robles (1997); Harberger (1998).]

When one gets down to the firm level, the two-deflator method is in its element. Rarely do we have decent time series on the price index of a firm's output. Treating many firms in one industry, one might then give all of them the same price index—that of the industry. At that point the distribution of adjusted TFP residuals among the firms would end up differing from the original two-deflator distribution only by a constant. When expressed in percentage terms we would have $(R_j^*/y_j^*) = g_{yj}^* - s_{kj} g_{kj}^* - s_{\ell j} g_{\ell^* j}$ for each firm j without adjustment, while *with adjustment* we would have R_j^*/y_j = the same expression minus $g_{\bar{p}}$, the rate of growth of the *industry's* relative price index (the same for all firms).

So in the great bulk of cases one ends up with something quite like the two-deflator method when working at the individual firm level. The consolation is that the residual terms of individual firms, calculated for the whole economy, add up to a residual term for the aggregate—in the sense that outputs sum to GDP, the L_j^* sum to L^* for the economy, the K_j^* to K^* for the economy, etc. For more details on methodology, see Harberger (1998).

REFERENCES

Abramovitz, Moses. "Economics of Growth," in Bernard F. Haley, ed., *A survey of contemporary economics*. Homewood, IL: Richard D. Irwin (for the American Economic Association), 1952, pp. 132–78.

______. "Resources and Output Trends in the United States Since 1870." *American Economic Review*, May 1956 (*Papers and Proceedings*), *46*(2), pp. 5–23.

Aghion, Philippe and Howitt, Peter. "A Model of Growth Through Creative Destruction." *Econometrica*, March 1992, *60*(2), pp. 323–51.

Barro, Robert J. "Determinants of Economic Growth: A Cross-Country Empirical Study." Harvard Institute for International Development Discussion Paper No. 57[illegible], April 1997.

Barro, Robert J. and Sala-i-Martin, Xavier. *Economic growth*. New York: McGraw-Hill, 1994.

Beyer, Harald. "Sources of Economic Growth: Cross-Country Comparisons." Discussion paper, University of California, Los Angeles, 1996.

Boskin, Michael J., ed. *Capital formation and economic growth*. Stanford, CA: Hoover Institution, 1998 (forthcoming).

Denison, Edward F. *Why growth rates differ: Post-war experience in nine western countries*. Washington, DC: Brookings Institution, 1967.

DeSoto, Hernando. *The other path: The invisible revolution in the third world.* London: Tauris, 1989.

Easterly, William. "When Is Stabilization Expansionary? Evidence from High Inflations." *Economic Policy: A European Forum,* April 1996, (22), pp. 65–98.

Edwards, Sebastian. "Openness, Trade Liberalization and Growth in Developing Countries." *Journal of Economic Literature*, September 1993.

Fabricant, Solomon. *Economic progress and economic change*. Thirty-Fourth Annual Report of the National Bureau of Economic Research. New York: NBER, 1954.

Fischer, Stanley. "The Role of Macroeconomic Factors in Economic Growth." *Journal of Monetary Economics*, December 1993, *32*(3), pp. 485–512.

Griliches, Zvi. "Measuring Inputs in Agriculture: A Critical Survey." *Journal of Farm Economics*, December 1960, *42*(5), pp. 1411–27.

———. "The Sources of Measured Productivity Growth: United States Agriculture, 1940–60." *Journal of Political Economy,* August 1963, *71*(4), pp. 331–46.

———. "Productivity, R&D, and the Data Constraint." *American Economic Review*, March 1994, *84*(1), pp. 1–23.

Grossman, Gene M. and Helpman, Elhanan. *Innovation and growth in the global economy*. Cambridge, MA: MIT Press, 1991.

———. "Endogenous Innovation in the Theory of Growth." *Journal of Economic Perspectives*, Winter 1994, *8*(1), pp. 23–44.

Harberger, Arnold C. "Perspectives on Capital and Technology in Less Developed Countries," in M. J. Artis and A. R. Nobay, eds., *Contemporary economic analysis*. London: Croom Helm, 1978, pp. 42–72.

———. "In Step and out of Step with the World Inflation: A Summary History of Countries, 1952–76," in M. J. Flanders and A. Razin, eds., *Development in an inflationary world*. New York: Academic Press, 1981, pp. 35–46.

———. "Economic Policy and Economic Growth," in Arnold C. Harberger, ed., *World economic growth*. San Francisco, CA: ICS Press, 1984, pp. 427–66.

———. "Reflections on the Growth Process." Working paper, World Bank, 1990.

———. "Studying the Growth Process: A Primer," in Michael J. Boskin, ed., *Capital formation and economic growth*. Stanford, CA: Hoover Institution, 1998 (forthcoming).

Jorgenson, Dale W. *Productivity*, Vols. 1 and 2. Cambridge, MA: MIT Press, 1995.

Jorgenson, Dale W.; Gollop, Frank M. and Fraumeni, Barbara M. *Productivity and U.S. economic growth*. Cambridge, MA: Harvard University Press, 1987.

Jorgenson, Dale W. and Griliches, Zvi. "The Explanation of Productivity Change." *Review of Economic Studies*, July 1967, *34*(3), pp. 249–80.

Kendrick, John W. *Postwar productivity trends in the United States.* New York: Columbia University Press, 1973.

———. *The formation and stocks of total capital.* New York: National Bureau of Economic Research, 1976.

———. *Understanding productivity: An introduction to the dynamics of productivity change.* Baltimore, MD: Johns Hopkins University Press, 1977.

Kendrick, John W. and Grossman, Elliot S. *Productivity in the United States: Trends and cycles*. Baltimore, MD: Johns Hopkins University Press, 1980.

Krueger, Anne O. "Trade Policies in Developing Countries," in Ronald W. Jones and Peter A. Kenen, eds., *Handbook of international economics*, Vol. 1. Amsterdam: North-Holland, 1985.

———. "Trade Policy and Economic Development: How We Learn." *American Economic Review*, March 1997, *87*(1), pp. 1–22.

Lucas, Robert E., Jr. "On the Mechanics of Economic Development." *Journal of Monetary Economics*, July 1988, *22*(1), pp. 2–42.

———. "Making a Miracle." *Econometrica*, March 1993, *61*(2), pp. 251–72.

Mairesse, Jacques and Griliches, Zvi. "Heterogeneity in Panel Data: Are There Stable Production Functions?" in Paul Champsaur et al., eds., *Essays in honor of Edmond Malinvaud, volume 3: Emprical Economics.* Cambridge, MA: MIT Press, 1993, pp. 192–231.

Mansfield, Edwin. *Innovation, technology and the economy: Selected essays of Edwin Mansfield.* Aldershot, U.K.: Elgar, 1995.

North, Douglass C. *Institutions, institutional change and economic performance*. Cambridge, MA: Cambridge University Press, 1990.

Olson, Mancur, Jr. "Big Bills Left on the Sidewalk: Why Some Nations Are Rich and Others Poor." *Journal of Economic Perspectives,* Spring 1996, *10*(2), pp. 3–24.

Robles, Edgar. "An Exploration into the Sources and Causes of Economic Growth in the United States and Fourteen Latin American Countries." Ph.D. dissertation, University of California, Los Angeles, 1997.

Romer, Paul. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, October 1986, *94*(5), pp. 1002–37.

———. "Endogenous Technological Change." *Journal of Political Economy*, October 1990, *98*(5), pp. S71–S102.

———. "The Origins of Endogenous Growth." *Journal of Economic Perspectives*, Winter 1994a, *8*(1), pp. 3–22.

———. "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions." *Journal of Development Economics*, February 1994b, *43*(1), pp. 5–38.

Schmookler, Jacob. *Invention and economic growth.* Cambridge, MA: Harvard University Press, 1966.

Schumpeter, Joseph A. *The theory of economic development*. Cambridge, MA: Harvard University Press, 1934.

Solow, Robert M. "Technical Change and the Aggregate Production Function." *Review of Economics and Statistics*, August 1957, *39*(3), pp. 312–20.

Torre, Leonardo. "Concentration Patterns of the Contribution of TFP to Output Growth: Evidence From the Mexican Manufacturing Sector." Paper presented at the meeting of the Western Economic Association, 1997.

Williamson, John, ed. *Latin American adjustment*. Washington, DC: Institute for International Economics, 1990.

Competition Between Private and Public Schools, Vouchers, and Peer-Group Effects

By Dennis Epple and Richard E. Romano*

A theoretical and computational model with tax-financed, tuition-free public schools and competitive, tuition-financed private schools is developed. Students differ by ability and income. Achievement depends on own ability and on peers' abilities. Equilibrium has a strict hierarchy of school qualities and two-dimensional student sorting with stratification by ability and income. In private schools, high-ability, low-income students receive tuition discounts, while low-ability, high-income students pay tuition premia. Tuition vouchers increase the relative size of the private sector and the extent of student sorting, and benefit high-ability students relative to low-ability students. (*JEL* H42, I28)

Discontent in the United States with the pri-ary and secondary educational system has come the norm. The decline in SAT scores the 1970's, embarrassing international com-risons of student achievement, slow growth productivity measures, and increasing dis-rity in earnings all call into question the ality of the educational system.[1] Education policy figured prominently in recent presidential elections. The debate has centered on issues of school choice, including voucher systems (Karen De Witt, 1992). Typical voucher proposals provide students attending private schools a tax-financed, school-redeemable voucher of fixed amount toward (or possibly covering) tuition. Although a 1993 California referendum for vouchers was defeated, policy change at state and local levels abounds, as does change in the private educational sector. The state of Minnesota and school districts in 30 states allow residents to choose the public school their children attend.[2] The city of Milwaukee introduced a voucher system in the 1989–1990 school year. A number of private school and private–public school initiatives are developing (see e.g., John F. Witte et al., 1993; Steve Forbes, 1994; Steven Glazerman and Robert H. Meyer, 1994; Joe Nathan, 1994; *Newsweek,* 1994; *Wall Street Journal,* 1994; Steven Baker, 1995; Jay P. Green et al., 1996). Educational reform emphasizing increased school competition with an increased

* Epple: Graduate School of Industrial Administration, rnegie Mellon University, Pittsburgh, PA 15213; mano: Department of Economics, University of Flor-, Gainesville, FL 32611. We greatly appreciate the mments of Linda Argote, Richard Arnott, Lawrence nny, Tracy Lewis, David Sappington, Suzanne tchmer, and three anonymous referees, in addition to rkshop participants at Carnegie Mellon University, rida State University, Indiana University, Northwest- University, Princeton University, the University of icago, the University of Colorado, the University of rida, the University of Illinois, the University of Kan-, the University of Virginia, Yale University, the 1993 blic Choice meetings, and the 1994 American Eco-nic Association meetings. We thank the National Sci-e Foundation, and Romano thanks the Public Policy earch Center at the University of Florida for financial port. Epple acknowledges the support of Northwestern versity, where some of this research was conducted. y errors are ours.

[1] The provocatively titled report of the National Com-sion on Excellence in Education (1983), *A Nation at* , details the decline of performance of U.S. students e 1970's. More recent data can be found in Daniel M. etz (1987). Modest gains in performance on standard- achievement tests, followed by a leveling off, well w peak scores of the early 1960's, characterizes the 1980's and 1990's.

[2] Public funding of nonsecular schools and considerable freedom of school choice has been practiced for years in England (Daphne Johnson, 1990) and much of Canada (Nick Kach and Kas Mazurek, 1986). These choice systems support horizontal differentiation in schooling and safeguards exist to limit vertical (quality) differentiation. Our analysis is concerned primarily with the effects of a voucher system on vertical differentiation.

role of the private sector is at the forefront of the policy debate and recent policy initiatives.

The modern economic case for vouchers and increased educational choice was made by Milton Friedman (1962). The academic educational and political-science professions have since considered the pros and cons of voucher systems and educational choice (John E. Coons and Stephen D. Sugarman, 1978; Myron Liberman, 1989; John Chubb and Terry Moe, 1990). Economic analysis of the interaction between public and private schools, and of related policy instruments like vouchers, is only beginning to emerge. This paper continues the study of the "market" for education by developing a model that focuses on the interaction between the public and private educational sectors and also examines the consequences of vouchers. We describe the equilibrium characteristics of the market for education with an open-enrollment public sector and a competitive private sector.

Our model embodies two key elements of the educational process. First, students differ in their abilities. Higher ability is assumed to increase a student's educational achievement *and* that of peers in the school attended. Second, households differ in their incomes, with higher income increasing the demand for educational achievement. A student in our model is then characterized by an ability and a household income, a draw from a continuous bivariate distribution. A school's quality is determined by the mean ability of the student body, reflecting the model's peer-group effect. We characterize the equilibrium distribution of student types across public and private schools and examine the tuition structure of private schools, assuming that student types are verifiable. We develop a theoretical and computational model in parallel, with the latter calibrated to existing estimates of parameter values. Equilibria are simulated for a range of voucher values.

Key characteristics of an equilibrium are the following. A hierarchy of school qualities will be present, with the set of (homogeneous) public schools having the lowest-ability peer group and a strict ability-group ranking of private schools. The equilibrium student bodies of schools correspond to a partition of the ability–income-type space of students with stratification by income and, in many cases, stratification by ability. As Figure 1 from our computational model illustrates, type space is then carved into diagonal slices with each higher slice making up a private school's student body and with the bottom slice comprising the public sector.

The normality of demand for a good peer group leads relatively high-income students to cross subsidize the schooling of relatively high-ability students, producing the latter partition. Private schools attract high-ability low-income students by offering them tuition discounts, sometimes fellowships. Even with free entry, schools price discriminate by income against students who are not on the margin between switching schools. The equilibrium differentiation of schools and economies of scale in education preclude perfect competition for every type of student. Nevertheless, this price discrimination does not disrupt the internalization of the peer-group externality by private schools. An equilibrium without a public sector is Pareto efficient given the equilibrium number of schools. Because free public schools do not price the peer-group externality, an equilibrium with public schools is Pareto inefficient.

In the computational model, we employ a Cobb-Douglas specification of utility and educational achievement which incorporates the peer-group effect. The parameters are calibrated to U.S. data from various sources. We compute approximate equilibria for voucher values ranging from $0 to $4,200 per student ($4,222 equals the expenditure per student in public schools in 1988).[3] With no vouchers the predicted percentage of students in the public sector is 90 percent (the actual value for the United States is 88 percent). As the voucher is increased, the size of and mean ability in the public sector decrease. With a $2,000 voucher, for example, the percentage of students remaining in the public sector equals 70 percent, and the mean ability declines by 15.8 percent.

[3] The integer number of private schools in our model precludes existence of competitive equilibrium except in special cases. This integer problem and our approximation approach are discussed later in the paper.

The entry of private schools and consequent more efficient sorting of students across schools caused by vouchers increases average welfare (and achievement) only a little in our computational model, while having larger distributional effects. As we discuss in detail later, the magnitude of the aggregate effect depends on the extent of complementarity of peer ability and own ability in the educational production function. There is little empirical evidence to guide assessment of the extent of such complementarity. The voucher increases the premium to ability in private schools. The largest proportionate gains from the voucher then accrue to low-income, high-ability students. For example, a household with income $10,000 and student with ability at the 95th percentile has a welfare gain of about 7.5 percent of income from a $2,000 voucher. Students of low income and low ability who remain in the public sector when a $2,000 dollar voucher is available experience small welfare losses but make up a majority. It bears emphasizing that our model takes public and private schools to be equally effective providers of education, however. Some argue that private schools are more productive and that the competitive effect of a voucher program will increase public-school effectiveness.[4] For example, Hoxby (1996) concludes from her empirical investigation that competition-induced performance improvement would increase public-school achievement by more than enough to offset losses of the magnitude that emerge due to reduced peer quality in our computational model. Our analysis delineates the allocative effects of vouchers and demonstrates a potential for significant redistribution.

A theoretical-economics literature on education is beginning to emerge. Charles A. M. de Bartolomé (1990) develops a two-neighborhood model of the provision of public educational inputs (quality) with two ability types and peer-group externalities. He shows that the voting/locational equilibrium is inefficient because the median voter does not internalize the consequences of migration on peer groups in choosing the input level. No independent income variability characterizes students in his model. Raquel Fernandez and Richard Rogerson (1996) introduce income differences in a two-neighborhood model of the provision of inputs but abstract from peer-group effects. They examine the effects of redistributive policies and direct controls on inputs. Neither model has a private sector. Our analysis is differentiated by its consideration of a private sector and its two-dimensional, continuous type space. In a normative analysis of student groupings in the presence of peer-group effects, Richard Arnott and John Rowse (1987) show how a social planner would maximize the sum of achievements in allocating students of various abilities across classrooms. We analyze equilibrium outcomes, and most of our analysis is positive.

Joseph E. Stiglitz (1974), Norman J. Ireland (1990), Ben Eden (1992), Charles F. Manski (1992), Michael Rothschild and Lawrence J. White (1995), Epple and Romano (1996), and Gerhard Glomm and B. Ravikumar (1998) consider the consequences of a private sector for education. Stiglitz, Glomm and Ravikumar, and Epple and Romano are concerned with the existence and properties of voting equilibria over tax-financed, public-school expenditure in the presence of a private alternative. Ireland analyzes the effects of vouchers on utilities and the quality of the public alternative, taking the tax rate as exogenous. Individuals differ only by income, and the private alternative can be purchased continuously in all these analyses. Hence, the private sector is relatively passive, and issues of financial aid and differences in

[4] Caroline M. Hoxby (1994, 1996) provides evidence that private-school competition increases public-school effectiveness. William N. Evans and Robert M. Schwab (1995) find that Catholic private schools are more effective in inducing students to complete high school and also attend college. These studies take on the challenge of finding instruments that predict well private-school attendance while being independent of unobserved determinants of educational achievement. Controversy exists concerning the quality of the instruments used. See Thomas J. Kane (1996) for a discussion of Hoxby's methodology. David N. Figlio and Joe A. Stone (1997) employ a different set of instruments than Evans and Schwab and find that, at current input levels, religious private schools are less effective than public schools in producing achievement on standardized exams in math and science (but nonreligious private schools are more effective). See Kane (1996) and Figlio and Stone (1997) for references to other studies.

student ability across schools do not arise. Our model is distinguished by having differences in ability and related peer-group effects, and by providing an active role for private-sector schools.

Eden (1992) analyzes vouchers in a purely private market system of provision of education having two ability types and peer-group effects. A voucher equal to the difference between the social and private benefit of education to each ability type is shown to induce socially optimal provision of education. Key differences in our analysis include our consideration of the interaction between the public and private sectors, our exploration of the implications of continuous differences in ability and income, and our attention to positive issues. Manski (1992) pursues a computational analysis of vouchers that also considers peer-group effects among other aspects of education (especially various objectives of public-school decision makers). Our models differ in a number of ways. Most importantly, we permit private schools to discriminate in their tuition policies, with many consequences. Rothschild and White (1995) analyze a competitive model with consumers also inputs to production (a peer-group effect), using higher education as their primary example. We share a concern for market pricing in the presence of an externality. Differences in our model, among others, are the presence of a public sector, a more detailed specification of peer effects and demand for education, and student variation in both ability and household income. Our attention to the implications for pricing, profitability, and school qualities of a peer-group effect deriving from student abilities, the allocation of students according to ability and household income and the related distribution of educational benefits, and the effects of vouchers are not concerns in Rothschild and White.

Private schools are cases of clubs with non-anonymous crowding due to the ability-dependent externality and schools' power to price it. Suzanne Scotchmer (1994) provides an excellent synthesis of this literature. We follow this literature in our competitive specification of private schools as further discussed below.

The next section presents the model. Section II develops the theoretical results. The computational results comprise Section III. Concluding remarks follow. An Appendix contains some of the detail.

I. The Model

Household income is denoted y, and each household has a student of ability b. The joint marginal distribution of ability and income in the population is denoted $f(b, y)$ and is assumed to be continuous and positive on its support, $S \equiv (0, b_{\max}] \times (0, y_{\max}]$. All students attend a school since we assume that free public schooling is preferred to no schooling. The household decision maker's utility function, $U(\cdot)$, is increasing in numeraire consumption and the educational achievement of the household's student, and it is continuous and twice differentiable in both arguments. Achievement, $a = a(\theta, b)$, is a continuous and increasing function of the student's ability and the mean ability of the student body in the school attended, θ.[5] Let y_t denote after-tax income and

[5] The influence of ability on own educational achievement is well documented and not controversial. Eric Hanushek (1986) provides an excellent survey. In the economics literature, Anita A. Summers and Barbara L. Wolfe (1977) and Vernon Henderson et al. (1978) find significant peer-group effects. Evans et al. (1992) adjust for selection bias in the formation of peer groups and show that it eliminates the significance of the peer group in explaining teenage pregnancy and dropping out of school. They are careful to point out that their results should not be interpreted as suggesting that peer-group effects do not exist, but as demonstrating that scientific proof of these effects is inadequate. Note, too, that their work supports the notion that peer-group variables enter the utility function since a selection process does take place. The psychology literature on peer-group effects in education also contains some controversy. In their survey paper, Richard L. Moreland and John M. Levine (1992) conclude:

> The fact that good students benefit from ability grouping, whereas poor students are harmed by it, suggests that the mean level of ability among classmates, as well as variability in their ability levels, could be an important factor. The results from several recent studies . . . support this notion.

This squares with our reading of the literature (Summers and Wolfe, 1977; Henderson et al., 1978; Chen-Lin Kulik and James A. Kulik, 1982, 1984; Aage B. Sorensen, 198

tuition expenditure, the latter equal to zero if a public school is attended. Thus, $U = U(y_t - p, a(\theta, b))$, with U_1, U_2, a_1, and a_2 all positive. The achievement function captures the peer-group effect in our model, discussed further below. To maintain simplicity and highlight the role of peer groups, a school's quality is determined exclusively by the mean ability of its peer group.[6] In ongoing work we are extending the model to include variation in educational inputs.

U is also assumed to satisfy everywhere the "single-crossing" condition (SCI):

$$(1) \qquad \partial\left(\frac{\partial U/\partial\theta}{\partial U/\partial y_t}\right) \Big/ \partial y_t > 0.$$

Hence, for students of the same ability, any indifference curve in the (θ, p)-plane of a higher-income household cuts any indifference curve of a lower-income household from below. This condition corresponds to an income elasticity of demand for educational quality that is positive at all qualities for all types.[7] One set of sufficient conditions on U for SCI is $U_{11} \leq 0$ and $U_{12} \geq 0$, with at least one having strict inequality.[8]

Preferences for school quality might also depend on ability. We say preferences satisfy *weak single crossing in ability* if

$$\partial\left(\frac{\partial U/\partial\theta}{\partial U/\partial y_t}\right) \Big/ \partial b \geq 0$$

which implies a weakly positive ability elasticity of demand for quality. However, because the pertinent empirical evidence is mixed and scarce, we postpone restricting preferences in this regard until necessary.[9] In our computational model and to illustrate our more general theoretical results, we adopt a Cobb-Douglas specification of the utility function:

$$(2) \qquad U = (y_t - p)a(\theta, b)$$

$$a(\theta, b) = \theta^{\gamma} b^{\beta}$$

$$\beta > 0 \quad \gamma > 0.$$

While (2) satisfies SCI, it embodies the "neutral" assumption of zero ability elasticity of demand:

$$\partial\left(\frac{\partial U/\partial\theta}{\partial U/\partial y_t}\right) \Big/ \partial b = 0.$$

Our computational results are not driven by own-ability effects on the demand for education. Keep in mind, too, that the theoretical results do not assume specification (2).

A school's costs depend only on the number of students it enrolls, since inputs vary only with size. All schools, public and private, have the simple cost function:

$$(3) \qquad C(k) = V(k) + F$$

$$V' > 0 \quad V'' > 0$$

Sorensen and Maureen T. Hallinan, 1986; Adam Gamoran and Mark Berends, 1987; Jennie Oakes, 1987; Gamoran, 1992). However, there are alternative interpretations (Robert E. Slavin, 1987, 1990).

[6] For simplicity, the possibility that dispersion in peer ability also affects achievement is not built into our model. Roland Benabou (1996b) explores the consequences for economic growth of dispersion in human capital.

[7] We believe this to be uncontroversial. While we know of no empirical studies that use direct measures of educational quality, a substantial empirical literature on the demand for educational expenditure exists. Although considerable diversity in magnitudes of estimates of the income elasticity of demand for educational spending are present, estimates using a variety of approaches find the income elasticity to be positive (Daniel Rubinfeld and Perry Shapiro, 1989).

[8] Households may consider education a consumption good, an investment good, or a combination of the two. Our formulation can be interpreted to accommodate any of these motives. However, for households not subject to borrowing constraints, a pure investment motive would imply a zero income elasticity of demand. For such households, this in turn would imply that the SCI condition in (1) would be only weakly satisfied. In light of the empirical evidence suggesting the income elasticity to be positive, we conserve space in the development that follows by assuming that SCI is strict for all households.

[9] Henderson et al. (1978) find no interaction between own ability and the benefits to an improved peer group, corresponding to $\partial^2 U/\partial\theta\partial b = 0$ in our model. Summers and Wolfe (1977) find some support for higher peer-group benefits to lower-ability students, that is, $\partial^2 U/\partial\theta\partial b < 0$. Thus the literature provides limited evidence from which to draw conclusions.

where k denotes the number of attending students. Technical differences among schools are not an element of our model (for simplicity). Hence, vouchers cannot drive technically inefficient schools from the market, an effect predicted by some proponents of vouchers (see footnote 4). Let k^* denote the "efficient scale,"

$$k^* \equiv \mathrm{ARGMIN}[C(k)/k]. \tag{4}$$

The presumption of some economies of scale in education is realistic (Lawrence Kenny, 1982) and important. Otherwise, the private market would produce an infinite number of schools containing infinitely refined peer groups. Our model's equilibrium will be consistent with the fact that the number of types of students greatly exceeds the number of schools.

Public-sector schools offer free admission to all students. This open-enrollment policy leads to homogeneous public schools in equilibrium because we assume no frictions in public-school choice are present. Without equalization of θ's in public-sector schools, students would migrate to higher-θ schools to reap the benefits of a better peer group. With equalized θ's, no incentives for switching schools within the public sector remain. We study the alternative of neighborhood school systems that impose residence requirements in Epple and Romano (1995).

Since all public schools will have the same θ, one can think of the public sector as consisting of one (possibly large) school. Public-sector schooling is financed by a proportional income tax, t, paid by all households, whether or not the household's child attends school in the public sector. Thus, $y_t = (1 - t)y$. The public sector chooses the (integer) number of schools and their sizes to minimize the total cost of providing schooling subject to (3). The tax rate adjusts to balance the budget. Because households are atomistic, there is no tax consequence to a household's decision about school attendance. The public finance of schooling can then be largely suppressed in the analysis until the consideration of vouchers. The public sector is passive in this model for simplicity. Public-sector schools do not segment students by ability (track), increase educational inputs to compete more effectively with the private sector, or behave strategically in any way. More realistic alternatives are important topics for research, some of which are discussed in the final section.

Private-sector schools maximize profits, and there is free entry and exit.[10] Modeling private schools as choosing an admission policy and tuition policy is convenient and involves no loss of generality. Student types are observable, implying that tuition and admission can be conditioned on ability and income as competition permits.[11] Private schools are an example of clubs with "non-anonymous crowding" (Scotchmer and Myrna H. Wooders, 1987; Scotchmer, 1997) because of the peer-group effect, and we model private-school behavior following the literature on *competitive* club economies. In particular, private schools maximize profits as *utility takers* (see Scotchmer, 1994), a generalization of price-taking when consumers (types) and products differ. Private schools believe they can attract any student type by offering admission at a tuition yielding at least the maximum utility the student could obtain elsewhere.

Let an i subscript, $i = 1, 2, \ldots, n$, indicate a value for the ith private school. A zero subscript does the same for "the" public school. Let $p_i(b, y)$ denote the tuition necessary to enter school i, with $p_0(b, y) = 0\ \forall\ (b, y)$. Let $\alpha_i(b, y) \in [0, 1]$ denote the proportion of type (b, y) in the population that school i admits

[10] Consideration of alternative objective functions to profit maximization is reasonable, especially given the significant proportion of nonprofit schools. Some private schools might, for example, pursue the objective of quality maximization. Quality maximization, like profit maximization, is a member of a set of objective functions that is utility independent in the sense that they place no weight on offering any student types higher utility than the student's (equilibrium) reservation utility. Our preliminary analysis of this issue suggests that equilibria where some private schools pursue objectives from this set other than profit maximization must also be competitive equilibria. Roughly, the failure of any school to maximize profits would permit entry by a profit-maximizing school.

[11] The notion is that abilities can be determined through testing, and required financial disclosures permit determination of household income. At least in the case of Cobb-Douglas utility, equation (2), students will have no incentive to underperform on exams, since tuition will be nonincreasing in ability in equilibrium (proved in Epple and Romano [1993]). Incentive compatibility in the reporting of income is more complex.

with any $\alpha_0(b, y) \in [0, 1]$ "optimal" for the public school as determined by the residual demand for public education. A private school's profit-maximization problem can be written as

$$(5) \quad \underset{\theta_i, k_i, p_i(b,y), \alpha_i(b,y)}{\text{MAX}} \quad \pi_i \equiv \iint_S [p_i(b, y)\alpha_i(b, y) \times f(b, y)\, db\, dy] - V(k_i) - F$$

subject to

$$(5a) \quad \alpha_i(b, y) \in [0, 1] \; \forall (b, y);$$

$$(5b) \quad U(y_t - p_i(b, y), a(\theta_i, b)) \geq \underset{j \in \{0,1,\dots,n \mid j \neq i;\ \alpha_j(b,y) > 0 \text{ is in the optimal set of } j\}}{\text{MAX}} U(y_t - p_j(b, y), a(\theta_j, b)) \quad \forall (b, y);$$

$$(5c) \quad k_i = \iint_S \alpha_i(b, y) f(b, y)\, db\, dy;$$

$$(5d) \quad \theta_i = \frac{1}{k_i} \iint_S b\alpha_i(b, y) f(b, y)\, db\, dy.$$

Constraints (5c) and (5d) define, respectively, the size of the school's student body and the mean ability. Constraint (5a) precludes a school from admitting a negative number of a type or more of a type than exits in the population.[12] Constraint (5b) imposes the utility-taking assumption. Students' alternatives are limited to schools where they are admitted. Students always have the option of attending the public school. It is innocuous to require that (5b) hold for all (b, y) as we have specified (i.e., including for nonadmitted students). Tuition charged to students for whom $\alpha_i(b, y) = 0$ is school i's only optimal choice (i.e., nonadmitted students) is irrelevant. Note, too, that tuition such that (5b) holds with strict equality will be optimal.

Private schools enter so long as they expect to make positive profits as utility takers. Because incumbent private schools maximize profits as utility takers, entry results if and only if $\pi_i > 0$ for some incumbent school. The public-sector/private-sector equilibrium is described by the following five conditions in addition to the government balanced-budget condition presented below in Section II, subsection C, for the more general case with vouchers.

Condition UM:

$$U^*(b, y) = \underset{i \in \{0,1,\dots,n \mid \alpha_i(b,y) > 0 \text{ is in the optimal set of } i\}}{\text{MAX}} U(y_t - p_i(b, y), a(\theta_i, b)) \quad \forall (b, y).$$

Condition ΠM:

$$[\theta_i, k_i, p_i(b, y), \alpha_i(b, y)] \text{ satisfy (5)}, \quad i = 1, 2, \dots, n.$$

Condition ZΠ:

$$\pi_i = 0 \qquad i = 1, 2, \dots, n.$$

Conditions PSP:

$$p_0(b, y) = 0 \; \forall (b, y)$$

$$\alpha_0(b, y) \in [0, 1] \; \forall (b, y)$$

$$k_0 = \iint_S \alpha_0(b, y) f(b, y)\, db\, dy$$

$$\theta_0 = \frac{1}{k_0} \iint_S b\alpha_0(b, y) f(b, y)\, db\, dy.$$

[12] One might object to the presumption that "competitive schools" recognize the limit to demand. The presumption is analogous to a monopolistically competitive firm's recognition of a limit on its demand curve. Dropping the presumption would lead to schools admitting infinite densities of some types. See Scotchmer (1994) for the analogue in the literature on club goods.

Condition MC:

$$\sum_{i=0}^{n} \alpha_i(b, y) = 1 \quad \forall (b, y).$$

Condition UM summarizes household utility maximization. Households choose a most-preferred private or public school, taking admission/tuition policies, school qualities, and taxes as given. Profit maximization of private schools (ΠM) and the public-sector policies (PSP) have been discussed. While the entry assumption above is formally part of the definition of equilibrium, it is convenient to substitute the implication that private schools must earn zero profits (ZΠ). The last condition is market clearance, which uses the simplifying assumption above that free public schooling is preferred to no schooling.

II. Theoretical Results

A. *Solution to the Private School's Problem*

Using UM, the first-order conditions for problem (5) can be written as follows:

$$\text{(6a)} \qquad U(y_t - p_i^*, a(\theta_i, b)) = U^*(b, y) \quad \forall (b, y);$$

$$\text{(6b)} \qquad \alpha_i(b, y) \begin{Bmatrix} = 0 \\ \in [0, 1] \\ = 1 \end{Bmatrix} \text{ as } p_i^*(b, y, \theta_i) \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} V'(k_i) + \eta_i(\theta_i - b) \quad \forall (b, y);$$

$$\text{(6c)} \qquad \eta_i = \frac{1}{k_i} \iint_S \left[\frac{\partial p_i^*(b, y, \theta_i)}{\partial \theta_i} \alpha_i(b, y) \times f(b, y)\, db\, dy \right].$$

Condition (6a) describes school i's optimal tuition function, $p_i^*(b, y, \theta_i)$ and is just (5b) with equality combined with the equilibrium condition UM; $p_i^*(\cdot)$ is student-type (b, y)'s reservation price for attending school of quality θ_i. Condition (6b) characterizes optimal admission policies.[13] The term $\eta_i(\theta_i - b)$ may be thought of as the marginal cost of admission operating via the peer-group externality in school i. From (6c), η_i [the Lagrangian multiplier on (5d)] equals the per-student revenue change in school i deriving from a change in θ_i. The appropriately scaled change in θ_i due to admitting student of ability b equals $(b - \theta_i)$; its negative is then multiplied by η_i to obtain the peer-externality cost. The peer cost of admitting students with ability below the school's mean is positive because the resulting quality decline dictates reduced tuition to all students, while the peer "cost" of admitting above-mean-ability students is negative. Let $MC_i(b) \equiv V'(k_i) + \eta_i(\theta_i - b)$, which we term *effective marginal cost.* Types with reservation prices below $MC_i(b)$ are not willing to pay enough to cover their effective marginal cost and are not admitted. The school admits all of a type that has a reservation price above effective marginal cost, and any $\alpha_i \in [0, 1]$ is optimal if $p_i^* = MC_i$.[14]

B. *Properties of Equilibrium*

We now turn to the properties of equilibrium, assuming one exists. Existence issues are discussed below. Heuristic arguments have been substituted for formal proofs where reasonable.

The first result concerns the qualities of schools.

PROPOSITION 1: *A strict hierarchy of school qualities results, with the public sector*

[13] Results (6b) and (6c) are found by substituting p_i^* from (6a) into (5), and then forming a Lagrangian function to take account of (5c) and (5d). Result (6b) is then derived by pointwise optimization over α_i while taking account of the constraint (5a).

[14] In the upper and lower lines of (6b), the solution for α_i is at a corner, and the first-order conditions are also sufficient for a local maximum. In the middle line of (6b), where $p_i^* = MC_i$ and any $\alpha_i(b, y) \in [0, 1]$ satisfies the first-order conditions, V'' sufficiently large implies local maximization.

having the lowest-ability peer group: $\theta_n > \theta_{n-1} > \cdots > \theta_1 > \theta_0$.

Formal proof is in the Appendix. Here an economic interpretation is provided. All private schools must be of higher peer quality than schools in the public sector. Otherwise, no students would be willing to pay to attend any private school.

Why must a strict hierarchy of private schools characterize equilibrium? If two private schools were of the same quality, then they would compete perfectly for students. Consequently, they would have the same effective marginal costs of admitting all types, and their tuitions (to all admitted students) would equal effective marginal costs. An opportunity to increase profits would exist by varying admissions/tuitions in such a way to either: (a) increase quality and admit a student body that values quality by more, or (b) decrease quality and admit a student body that values quality by less. In either case, the school differentiates itself in quality, at the same time attracting a student body that permits profitable price discrimination over the quality change.

We sketch the example of a profitable quality improvement, beginning with schools having identical student bodies (the proof shows that this is without loss of generality). Let one school admit the same number of (b_2, y_2) types as it expels of (b_1, y_1) types, where $b_2 > b_1$ and $y_2 > y_1$, implying an increase in θ but no change in production costs, $V(k) + F$. Further, choose the types (which is always feasible) such that $y_2 - y_1 > b_2 - b_1$ by enough that, using SCI, the (b_2, y_2)-types value increased quality by more than the (b_1, y_1)-types, even though their abilities differ. This permits the school to charge the newly admitted students tuitions higher than their effective marginal costs because they are selected to value quality increases by more than the expelled students. The profit increase occurs because the new student body values the quality increase by more than would the original student body; θ and η rise in the school. It would not increase profits to substitute students in such a way that θ rises without also changing the student body's average value of quality improvements, because tuitions equal effective marginal costs in the initially nondifferentiated schools.[15] This example assumes that a school substitutes students to increase quality, but alternative profitable substitutions exist that decrease quality, roughly, by also creating a lower-income student body.

In either case, the argument depends on SCI. It also identifies the model's force for "diagonal stratification" (see the examples in Figure 1). As developed more fully below, this stratification results because students having relatively high income and low ability within a school cross subsidize relatively low-income, high-ability students.

The strict hierarchy of Proposition 1 supports the equity-related concerns of some that private schools operate to the detriment of public schools by siphoning off higher-ability students. Whether a strict hierarchy is efficient is analyzed below. First we develop further the positive properties of equilibrium.

Proposition 2 describes equilibrium pricing, and Proposition 3 describes the resulting partition of types. Some definitions are useful. Let $\mathcal{A}_i \equiv \{(b, y) \in S \mid \alpha_i(b, y) > 0 \text{ is optimal}\}$ denote the *admission space* of school i, $i = 0, 1, \ldots, n$ (see Figure 1, for example). A locus of points $(b, y) \in \mathcal{A}_i \cap \mathcal{A}_j$, $i \neq j$, assuming it exists, is referred to as a *boundary locus* between i and j. (Boundary loci have zero measure in S, as proved in Epple and Romano [1993].) Since any household prefers free public schooling to no schooling, the entire type space S is partitioned into admission spaces. Last, to avoid tedious qualification of statements for public-sector schools, we specify that $\mathrm{MC}_0 \equiv 0$ for all (b, y). This notation is convenient since students see a zero cost of public education.

PROPOSITION 2: *(i) On a boundary locus between school i and j,* $p_i = \mathrm{MC}_i(b)$ *and* $p_j = \mathrm{MC}_j(b)$; *pricing on boundary loci is strictly according to ability in private schools. (ii)* $p_i(b, y) > \mathrm{MC}_i(b)$ *for off-boundary students who attend private school i; pricing off-*

[15] Mathematically, beginning with equal θ's, first-order effects on profits of varying admissions vanish, but the profit function is convex in some directions in $[\alpha(b, y), p(b, y)]$-space, allowing a profit increase.

boundary loci depends on income in private schools. (iii) Every student attends a school that would maximize utility if all schools instead set p_i equal to equilibrium MC_i *for all students. The allocation is as though effective marginal cost pricing prevails in private schools.*[16]

See Epple and Romano (1993) for proof. Competition between private schools that share a boundary locus forces prices to effective marginal costs for student-types on the locus. These students are indifferent to attending the schools sharing the locus. Private schools then have no power to price discriminate with respect to income on boundary loci. Prices are, however, adjusted to differing abilities because private schools internalize the peer-group effect. Tuition to private school i decreases with ability at rate η_i along its boundary loci, reflecting the value of peer-group improvements of the school's student body.

Moving inside a boundary locus in a private school's admission space, students' preferences change in such a way that they would strictly prefer the school attended if it practiced effective marginal-cost pricing. Part (ii) of Proposition 2 establishes that private schools exploit this by increasing price. These students are also indifferent between the private school attended or their best alternative by (6a), but this is a *result* of discriminatory pricing. Generally, then, price depends both on ability and income within admission spaces.[17]

Part (iii) of Proposition 2 follows because it is profitable for a private school to be sure to attract any student whose reservation price exceeds the school's effective marginal cost. The student allocation's link to effective marginal costs, and hence abilities, will be shown to be efficient (except for the public sector). The income-related price discrimination that occurs does not disrupt the allocation consistent with effective marginal-cost pricing; rather, it is purely redistributive.

While this income-related price discrimination is of the first degree (à la Pigou), its magnitude is limited by competition for students among the differentiated schools. Near a boundary in a school's admission space, a student's preference for the school attended would be slight under effective marginal-cost pricing, so that the admitting school can capture little rent. The number and sizes of private schools then determine their power to price discriminate over income. All private schools have student bodies less than k^* by a similar argument to that in more standard monopolistically competitive equilibria.[18] Here school i's marginal-revenue curve can be constructed by ordering from highest to lowest students' reservation prices minus peer costs [i.e., $p_i^* + \eta_i(b - \theta_i)$], and thus the associated *downward-sloping* average revenue curve may be derived. Zero profits then implies a scale below k^*. If we let k^* decline, then private schools become more numerous and less differentiated (have closer θ's), and income-related price discrimination declines.

Now consider the partition of types into schools. We say *stratification by income* (SBI) holds if, for any two households having students of the same ability, one household's choice of a higher-θ school implies it has weakly higher income than the other household. Analogously, *stratification by ability* (SBA) is present if, holding income fixed, the household that chooses a higher-θ school must have a student of weakly higher ability. The combination of SBI and SBA implies a diagonalized partition as, for example, in Figure 1.

PROPOSITION 3: *(i) SBI characterizes equilibrium. (ii) If preferences satisfy weak single crossing in ability (W-SCB) and* $\eta_1 \le$

[16] The statements regard the *equilibrium* effective marginal cost. Income effects would cause these costs to change if tuition equaled effective marginal cost for all students. This has distributional (but not efficiency) implications.

[17] While there are no published studies of the allocation of financial aid by income and ability among private elementary and secondary schools, there is evidence on the allocation of financial aid by colleges and universities. There the evidence is that both ability and family income are significant determinants of whether and how much financial aid is received (J. Brad Schwartz, 1986; Sandra R. Baum and Saul Schwartz, 1988; Charles T. Clotfelter, 1991).

[18] The points made here are proved in Epple and Romano (1993).

$\eta_2 \le \cdots \le \eta_n$, *equilibrium.*[19]

To confirm p
holds with stude
fering incomes
indifference cur
sloping. For the
any indifferenc
ence curve of
as if tuitions e
[part (iii) of P
between schoo
the (θ, p)-pla
$\mathrm{MC}_i(b)$) and (
be that MC_j(
chooses i. A
ment then app

Part (ii) is
we provide so
the demand
ability (e.g.,
fication) and
same discoun
ary loci (i.e.
Holding non
real income
due to tuition
SBA would t
plaining SBI
positive inc
counts to abi
and SBA, t
Figure 1. R
ability stud
low-income
private scho
strongly if
SCB holds
dition is ne
our other re
SBA. It ma
ditions to
boundary lo

[19] We thank
to investigate t
1989), which

[20] The alter
types are willi
the alternative

$\eta_2 \le \cdots \le \eta_n$, then SBA also characterizes equilibrium.[19]

To confirm part (i), consider two households with students of the same ability but differing incomes $y_r^2 > y_r^1$. In the (θ, p)-plane, indifference curves of a household are upward sloping. For the same ability, SCI implies that any indifference curve of y_r^2 cuts any indifference curve of y_r^1 from below. Allocations are as if tuitions equal effective marginal costs [part (iii) of Proposition 2]. Thus, the choice between schools i and j may be represented in the (θ, p)-plane as a choice between $(\theta_i, \mathrm{MC}_i(b))$ and $(\theta_j, \mathrm{MC}_j(b))$. If $\theta_j > \theta_i$, it must be that $\mathrm{MC}_j(b) > \mathrm{MC}_i(b)$ if either type chooses i. A standard single-crossing argument then applies to complete the proof.

Part (ii) is proved in the Appendix; here we provide some intuition. Assume first that the demand for quality is independent of ability (e.g., as in the Cobb-Douglas specification) and that all private schools give the same discount to ability along their boundary loci (i.e., schools' η's are the same). Holding nominal household income fixed, real income would rise with student ability due to tuition discounts at all private schools. SBA would then result by the same logic explaining SBI. Hence, the combination of a positive income elasticity (SCI) and discounts to ability alone would cause *both* SBI and SBA, the diagonalized partition as in Figure 1. Relatively high-income and low-ability students cross subsidize relatively low-income and high-ability students in private schools. The argument holds more strongly if the η's strictly ascend or if W-SCB holds strictly. However, neither condition is necessary for SBA, nor do any of our other results require these conditions or SBA. It may be possible absent these conditions to get cases having nonmonotonic boundary loci in the (b, y)-plane.[20]

We now turn to normative results which are quite intuitive. Again, see Epple and Romano (1993) for the formal analysis. Pareto efficiency requires: (i) a student allocation that internalizes the peer-group externality given the number of schools, and (ii) entry as long as aggregate household net willingness to pay for an allocation with one more school exceeds the change in all schools' costs. An equilibrium without a public sector would satisfy condition (i) but not condition (ii). Effective marginal cost includes the marginal value of the peer group externality, implying that $\mathrm{MC}_i(b)$ equals the social marginal cost of attendance at school i by a student of ability b. A purely private-school equilibrium then satisfies efficiency condition (i) by part (iii) of Proposition 2.

However, entry to the point of zero profits entails externalities so that efficient entry [condition (ii)] fails to hold in a fully private equilibrium. An entrant captures the full value of its product to the student body it admits but ignores utility changes of nonadmitted students and profit changes of other schools resulting from the reallocation.[21] Fixed costs,

[19] We thank an anonymous referee for encouraging us to investigate bid-rent functions (see e.g., Masahisa Fujita, 1989), which ultimately led to part (ii) of Proposition 3.

[20] The alternative to W-SCB implies that lower-ability types are willing to pay more for a better peer group, and the alternative to weakly ascending η's implies that lower-quality schools give bigger discounts to ability. Either would tend to work against pure ability stratification, though Proposition 1 implies that some degree of ability stratification would be present. It is desirable to demonstrate SBA without assuming ascending η's, since these values are endogenous. However, providing general primitive conditions for SBA independent of assumptions concerning the equilibrium η's is difficult, because their equilibrium values depend on the entire distribution of types in the population. For the Cobb-Douglas case and assuming independence of income and ability in the population, we (Epple and Romano, 1993) have shown SBA without assuming weakly ascending η's.

[21] The comparison of the equilibrium number of schools in a fully private equilibrium to the Pareto-efficient number entails a trade-off. The entrant ignores the lost revenues and cost savings to other schools from the students that it admits. Since almost every student attracted away from incumbent schools is inframarginal (i.e., tuition exceeds effective marginal cost), the net effect here of entry is negative, tending to cause too much entry. Opposing this is the entrant's failure to capture the full returns from increased variety of school qualities that results. Although the entrant fully price discriminates to the students it admits, it cannot tax other students for the adjustments in the incumbent schools' qualities. A net benefit to other students is likely to result because the incumbent schools will better accommodate preferences.

hence the finite size of an entrant, underlie the entry externalities as in many models of monopolistic competition.

Introduction of the free public sector implies deviations from both efficiency conditions. In general, the public sector displaces multiple differentiated private schools, substituting the equivalent of one "large" homogeneous school. This effective reduction in the number of schools is without attention to costs and benefits, generally implying a deviation from efficiency condition (ii).

Holding fixed the number of schools in the public–private equilibrium (and counting the public sector as one school), zero pricing of public schooling violates condition (i). By *just* reallocating students between the public sector and private school 1 near their shared boundary locus, Paretian gains are feasible. Reference to the upper panel of Figure 1 from our computational equilibrium may help clarify the argument. Gains would result from shifting into private school 1 relatively lower-ability students below but near the boundary locus, students for whom the marginal social cost in the public sector is positive. These students are nearly indifferent between the two schools when facing the social cost of attending the private school but a tuition (zero) below the social cost of attending the public school. Students near the boundary locus and attending the private school may also be of sufficiently high ability that the social "cost" of attending the public school is negative. Gains from shifting such students into the public sector are then also feasible. Such students exist in our computational model, the rough prescription being to rotate the boundary locus counterclockwise at the point of ability having zero social marginal cost in the public school. Collecting these results, we have the following proposition.

PROPOSITION 4: *(i) The allocation in a fully private equilibrium is (Pareto) efficient given the number of schools; the equilibrium number of schools is not, however, generally efficient. (ii) The public–private-sector equilibrium has neither an efficient number of schools, nor an efficient student allocation given the number of schools.*

When fixed costs of schooling are small, the departure from efficiency in a fully private equilibrium will be correspondingly small. Part (i) of Proposition 4 can then be interpreted as making a case for private schooling and the vouchers we study. However, we have some reservations concerning this efficiency result. First, we are sympathetic to the view of many that access to a quality education is a right and serves as a means to limit historical inequities. Second, longer-run externalities from education not considered by private schools, like reduced crime, may be present. For these reasons, we explore the consequences of vouchers on all types instead of just providing aggregate measures. A somewhat distinct concern arises because exact equilibrium exists only in special cases. The interpretation of the efficiency results in the approximate equilibrium we study is discussed in subsection D, below.

C. *Vouchers*

We examine tax-financed cash awards to all those attending private school.[22] No role for vouchers is present in the tuition-free public sector. Reformulate the model by everywhere adding the amount of the voucher, v, to y, for households that choose a private school. The government's budget constraint is:

$$(7)\quad \iint_S tyf(b, y)\,db\,dy = \iint_{\mathcal{A}_1 \cup \mathcal{A}_2 \cup \cdots \cup \mathcal{A}_n} vf(b, y)\,db\,dy + \hat{N}[F + V(\hat{k})],$$

This positive externality will tend to cause too little entry. We believe that too many or too few private schools are possible, but we have not proved this.

[22] Our model permits households to retain as income any excess of the voucher amount over the tuition paid to the private school of choice, thereby avoiding considerable complication.

where $\hat{N}$ and $\hat{k}$ denote, respectively, the cost-minimizing number and size of schools in the public sector that satisfy demand for public education. Vouchers lower the real price of private education and increase the demand for it. We examine the effects of vouchers in our computational model.

D. Existence of Equilibrium and an Approximate Equilibrium

As we discuss in the Appendix, exact equilibrium generally fails to exist due to the integer number of private schools. We examine an approximate equilibrium in our computational analysis. Our "epsilon-competitive equilibrium" requires that no (utility-taking) private school, incumbent or entrant, could increase profits by more than ε. Let $\pi_{\max}$ and $\pi_{\min}$ denote the maximum and minimum profits earned by incumbent schools [which maximize profits over $p(b, y)$ and $\alpha(b, y)$ locally], and replace $Z\Pi$ in the definition of equilibrium with

$$\text{MAX}[\pi_{\max}, \pi_{\max} - \pi_{\min}, -\pi_{\min}] \leq \varepsilon.$$

Here $\pi_{\max}$ equals the maximum potential profits to an entrant, and the maximum of the second two terms in the brackets equals the largest feasible profit increase by an incumbent school. The revised definition of equilibrium continues to require UM, PSP, MC, and local profit maximization by incumbent private schools [i.e., (6a)–(6c)]. Last, the number of private schools is the minimum number satisfying these requirements.

The epsilon equilibrium retains all the positive properties of an exact equilibrium except that private schools could gain ε in profits via global adjustments. The allocation of students in a fully private equilibrium would then continue to satisfy efficiency condition (i).

III. Computational Equilibrium Model and Illustrative Results

We develop a computational model to illustrate our results, to examine vouchers, and to explore issues for which comparative-static analysis may yield ambiguous results. We calibrate it to existing empirical evidence so that the results will provide at least suggestive evidence about the impact of policy interventions. However, scant empirical evidence exists on some important parameters of the model.

A. Specification and Calibration

We require specifications for the density of income and ability, the utility and achievement functions, and the cost function for education. We assume that $\begin{bmatrix} \ln(b) \\ \ln(y) \end{bmatrix}$ is distributed bivariate normal with mean $\begin{bmatrix} \mu_b \\ \mu_y \end{bmatrix}$ and covariance matrix

$$\begin{bmatrix} \sigma_b^2 & \rho\sigma_b\sigma_y \\ \rho\sigma_b\sigma_y & \sigma_y^2 \end{bmatrix}.$$

To calibrate the distribution of income, we use mean ($36,250) and median ($28,906) income for households from U.S. census data for 1989. With units of income in thousands of dollars, these imply that $\mu_y = 3.36$ and $\sigma_y = 0.68$.

We adopt specification (2) for the combined utility-achievement function. To calibrate the ability distribution we presume that educational achievement determines future earnings and that the benchmark economy is in a steady state. First, define normed achievement, a_N, as our achievement function raised to the power $1/\beta$ and multiplied by a constant, $a_N \equiv Ka^{1/\beta} = K\theta^{\gamma/\beta}b$.[23] Then, a student with ability b attending a school with a peer quality of θ is presumed to have future annual earnings (E) given by $\ln E = \ln a_N = \ln K + (\gamma/\beta)\ln\theta + \ln b$. This normalization is such that a percentage change in ability leads to the same percentage change in dollars earned. Henderson et al. (1978) report the change in achievement percentile that results from moving students from classes stratified by ability to mixed

[23] The constant of proportionality, K, is arbitrary. A convenient scaling is to set $K = E[b]^{-\gamma/\beta}$. This scaling has the property that, if all students in the population were to attend the same school (i.e., $\theta = E[b]$), then normed achievement would equal ability (i.e., $a_N = b$).

classes. An elasticity of achievement with respect to peer ability that is 30 percent as large as the elasticity with respect to own ability is representative of the results they report. We adopt the somewhat conservative value of $\gamma/\beta = 0.2$. To complete the calibration of the distribution of ability, we then assume that the observed household-income distribution is the income distribution that emerges in a steady-state equilibrium in our benchmark model.[24] This yields $\mu_b = 2.42$ and $\sigma_b = 0.61$. Thus, mean and median ability are 13.6 and 11.3, respectively, and the standard deviation of ability is 9.1.[25]

Gary Solon (1992) and David J. Zimmerman (1992) provide evidence on the correlation between father's income and son's income, and they both find that the best point estimate of this correlation is approximately 0.4. Intergenerational correlation in income arises from two sources: correlation between household income and student ability and, for given ability, correlation between income and quality of school attended. Hence, SBI suggests that the intergenerational correlation in incomes is an upper bound on the correlation between parent's income and child's ability. For purposes of sensitivity analysis, we then assume that $\rho \in [0, 0.4]$. For our benchmark case, we set $\rho = 0$, which is particularly convenient for our steady-state calibration of the model. This completes the calibration of $f(b, y)$.

We now complete the calibration of preferences. The Cobb-Douglas specification implies unitary price and income elasticities for school quality, θ. Given the absence of empirical evidence on the demand for quality, these are plausible focal values and are consistent with estimates of demand for school expenditure (see e.g., Theodore Bergstrom et al., 1982). This function also implies that the marginal rate of substitution between school quality and the numeraire is invariant to own ability. Empirical evidence is mixed about whether an improvement in peer group is more beneficial to high- or low-ability students. Hence, our model's assumption that the effect of peer group is not biased toward either high- or low-ability types seems an appropriate choice for a baseline model. If school quality could be purchased at a constant price per unit of quality, each household's expenditure on education relative to total expenditure on other goods would be $\gamma/(1+\gamma)$. The existing share of aggregate disposable personal income in the United States that is spent on education is approximately 0.056. Hence, we set $\gamma = 0.06$. Using $\gamma/\beta = 0.2$ from above, the calibrated utility–achievement function is then

$$U = (y_t - p)\theta^{0.06} b^{0.30}.$$

We chose a cost function that is quadratic in the percentage of students (or households) a school serves:

$$F + V(k) = 12 + 1{,}300k + 13{,}333k^2,$$

with parameters set as follows. Expenditure per student in public schools in 1988 was $4,222 (*Statistical Abstract*, 1991 p. 434) and there was $1/2$ student per household (*Statistical Abstract*, 1992 pp. 46, 139). We specified our benchmark case to have four private schools and chose parameter values such that average cost in equilibrium was approximately $4,200 per pupil.[26] Experimentation indicated that

[24] More precisely, we let the distribution of ability be lognormal, and we approximate by assuming that this generates a lognormal distribution of earnings. We set the first two moments of the distribution of earnings equal to the first two moments of the distribution of income. That is, we choose μ_b and σ_b such that our benchmark equilibrium has $E[a_N] = E[y]/m$ and $\mathrm{Var}[a_N] = \mathrm{Var}[y]/m^2$. The constant m is the ratio of employed workers per household to the number of students per household ($m = 2.6$ in 1990). The distribution of earnings will not be exactly lognormal because of the discrete difference in schools attended, even though the distribution of ability is presumed to be lognormal. If every student attended public school in the benchmark model, and hence faced the same θ, earnings would be exactly lognormal. The approximation is a good one because 90 percent of the students do attend public schools as we will see.

[25] Ability can be related to IQ. Using IQ $\sim \mathcal{N}(100, 256)$, one obtains $\ln b = -1.38 + 0.038(\mathrm{IQ})$. In our no-voucher steady state, this implies that a worker with an IQ of 100 has expected income of $22,074, and a 10-point increase in his IQ increases expected income to $32,510. See the discussion in what follows relating to Figure 6 and the calculation of expected steady-state income conditional on ability.

[26] We have presented the cost function in terms of the percentage of students served or, equivalently, the per-

equilibrium properties are not very sensitive to the benchmark number of schools, but rather are sensitive to the minimum of the average cost of schooling.

We set $\varepsilon = 4.2$. This is the minimum value sufficient to assure existence of epsilon equilibrium for voucher values varying from zero to \$4,200 per student.[27]

B. *Results*

For our benchmark equilibrium with no voucher, the public sector has 90 percent of the student population, and the four private schools combined serve the remainder. The actual U.S. percentage of students enrolled in public schools during this period equaled 88 percent. Increasing ρ from 0 to 0.4 reduces public-sector attendance to 88 percent. Effects on other variables of so changing ρ are also small, and the results that follow are for $\rho = 0$.

Other computational results are presented in Figures 1–6. The upper panel of Figure 1 presents the boundary loci and admission sets in type space, in addition to the equilibrium θ's and k's. Here and in some other figures, both absolute and percentile ability scales are provided for perspective. The lower panel displays the allocation for a voucher of \$1,800. The linear boundary loci derive from the Cobb-Douglas specification. For results we present, intersections of boundary loci, if any, occur very near the bounds of the support of type space. Such intersections are insignificant, but we have encountered cases having only a piecewise linear public–private boundary due to meetings of loci well interior to type space. In such cases, some (high-income) students in public-sector schools have other than the lowest-θ private school as their best alternative.

In addition to illustrating the strict hierarchy of schools, the data in Figure 1 show a negative correlation between school qualities and school size in the private sector. Those who attend lower-quality private schools face closer substitutes, flattening the derived average revenue curves (discussed above) of these schools and increasing their size in equilibrium. The negative correlation of school quality and school size is a testable prediction, and it seems plausible that the most elite schools are smallest.

Figure 2 shows effective marginal costs for the four private schools in equilibrium with no voucher, truncated at the maximum ability of students attending each school. For students on the admission boundaries of their schools, tuition rates equal effective marginal costs. While price discrimination by income as well as ability is practiced on the interior of schools' admission spaces, price is close to marginal cost for almost all students. Hence, as a first approximation, one can interpret the marginal-cost functions in Figure 2 as price functions. High-ability students are seen to receive a tuition discount (financial aid) in all schools. In addition, students of sufficiently high ability pay negative tuition (i.e., a tuition waiver plus a stipend) in the top three schools. For our calibration, the standard deviation of ability is about 9 units. The results imply tuition reductions of roughly \$2,350 and \$3,240

centage of households served, k. In terms of k, average cost reaches a minimum at \$2,100, with $k^* = 0.03$; \$2,100 can then be interpreted as the average cost *per household.* There are twice as many households as students in the United States. Letting s denote the *number* of students and substituting $s = k/2$, one sees that the minimum of the average cost *per student* is \$4,200. In our presentation, we focus on per-student measures of tuition and costs; the related per-household measures are simply half those of the per-student values.

[27] This value is about 7 percent of the cost of a school operating at a scale that minimizes cost per student. Relative to fixed cost, ε is approximately 35 percent. Of course, a minimal ε, however measured, is desirable. We have studied how the minimum ε varies as we vary efficient school scale, k^*, while holding average cost constant. We find that the requisite ε to support equilibrium varies approximately proportionately with k^* if fixed cost is varied proportionately with k^*. This suggests, as we would expect, that ε can be made as small as desired if k^* is made sufficiently small. We have also investigated increasing fixed cost while holding k^* and minimum average cost constant. This tends to reduce the ratio of ε to fixed cost but increases the absolute magnitude of ε required to sustain equilibrium. Our investigation reveals that substantive findings from the computational model are not sensitive to the choice of k^* or the relative magnitude of fixed to variable cost. Rather, the key aspect of costs is the value of average cost at the minimum, and as discussed in the text, this value is based on observed school costs. The problem with pursuing a calibration that further lowers ε is that it leads to a computationally unmanageable number of schools for large vouchers.

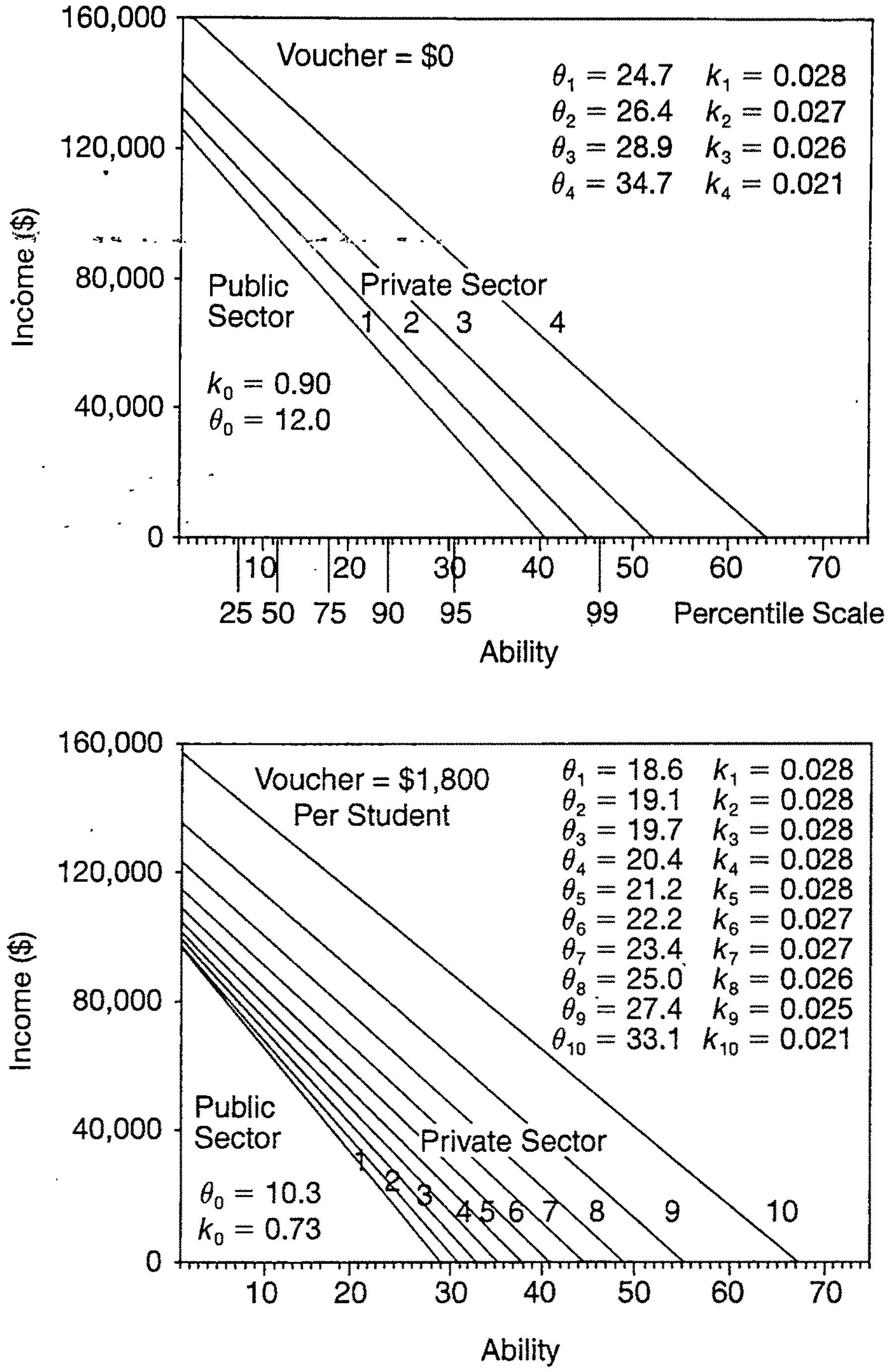


FIGURE 1. BOUNDARY LOCI WITHOUT AND WITH A $1,800 VOUCHER

for a one-standard-deviation increase in ability in schools 1 and 4, respectively.

The top panel of Figure 3 shows the ability distributions in the public-school sector and in the four private schools in the computed equilibrium. These distributions, all normalized to have an area of 1, illustrate the hierarchy among schools established in Proposition 1.

We simulated the effect of vouchers in amounts ranging from $0 to the minimum average cost of educating a student, $4,200. The effect of these vouchers on the distribution

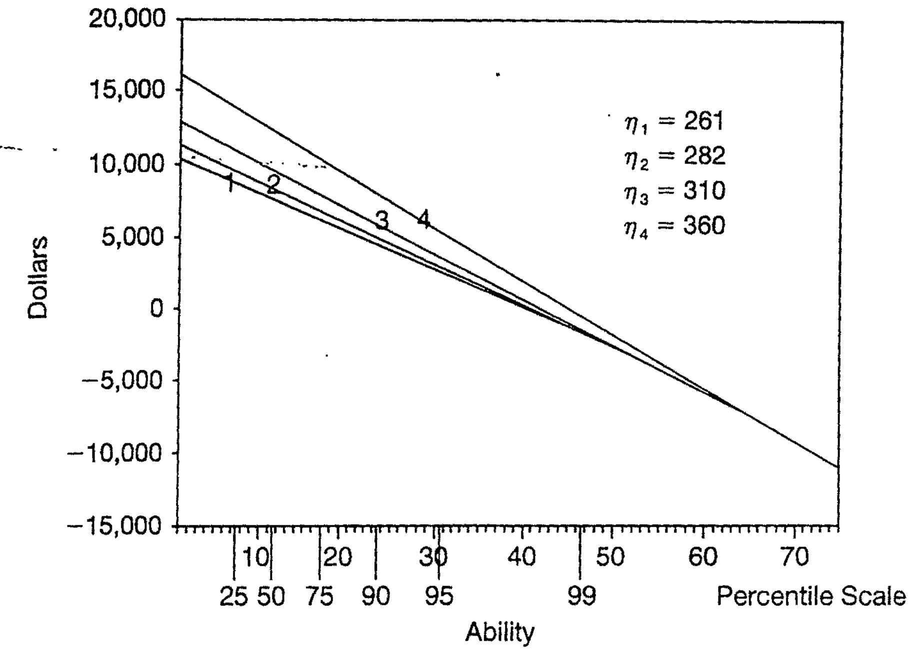


FIGURE 2. MARGINAL-COST FUNCTIONS WITH FOUR PRIVATE SCHOOLS

ability types attending public school is illustrated in the bottom panel of Figure 3. To illustrate the effect of the voucher program on public-school attendance of students of various ability levels, we have not normalized these distributions to have an area of 1. Thus, they represent the "numbers" of students at each ability level that are served by the public-school sector.

Figure 4 illustrates several ways in which equilibrium changes with the voucher. The upper panel shows the decline in the proportion of students attending public school as the voucher increases. The curve showing size during "transition" assumes the parental-income and student-ability distributions discussed above. When a voucher policy is introduced or varied (i.e., during transition), the income distribution of the parent generation will typically differ from the income distribution of the succeeding generation because the new voucher changes the distribution of earnings of a generation relative to that preceding. By contrast, in steady state, the income distribution is unchanging from generation to generation. To obtain the steady state for each value of the voucher, we hold the ability distribution fixed. We find the parent-income distribution that leads to an earnings distribution of the next generation that replicates the income distribution of the parents. The size of the public sector is virtually the same in transition and steady state (Figure 4), as are other variables that can be compared across the cases. To conserve space, other results presented are for the transition except when stated otherwise.

As expected, the size of the public-school sector declines to zero when the voucher nears the minimum average cost of providing education. The number of private schools at each voucher value is also shown in the top panel of Figure 4, illustrating the entry of new private schools as the voucher is increased. While the graph shows a continuous approximation, the number of schools is, of course, an integer at each voucher level in our computational model. We also calculated for each voucher level the percentage of households that favor that voucher level as compared to a voucher of zero. As the top panel of Figure 4 shows, support for a voucher program increases with the voucher but never reaches a majority. For example, only about 31 percent of the population benefits from a $2,000 voucher, and this includes the 10 percent who attend private

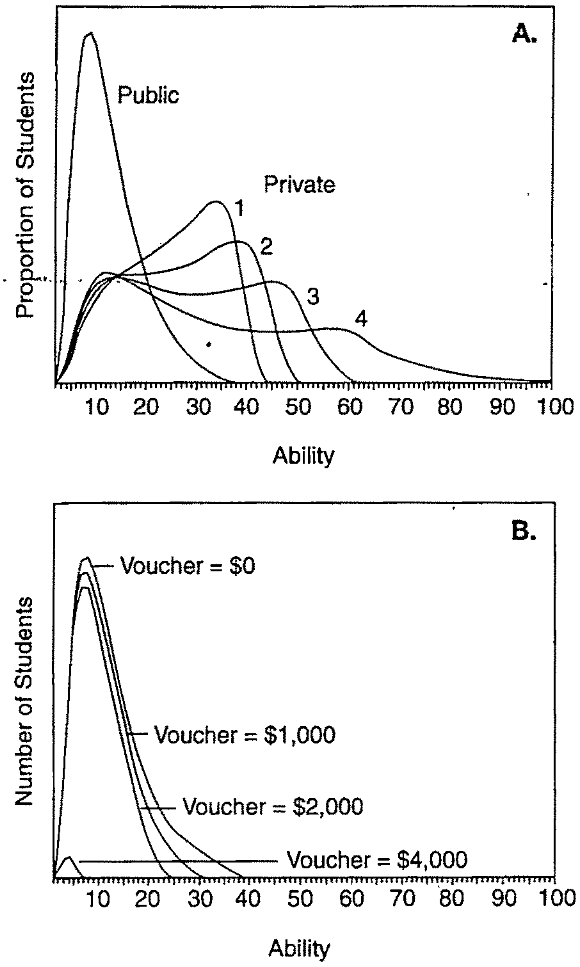


FIGURE 3. (A) ABILITY DISTRIBUTIONS IN PUBLIC AND PRIVATE SCHOOLS (VOUCHER = $0) AND (B) THE EFFECT OF VOUCHERS ON THE DISTRIBUTION OF ABILITIES OF STUDENTS ATTENDING PUBLIC SCHOOL

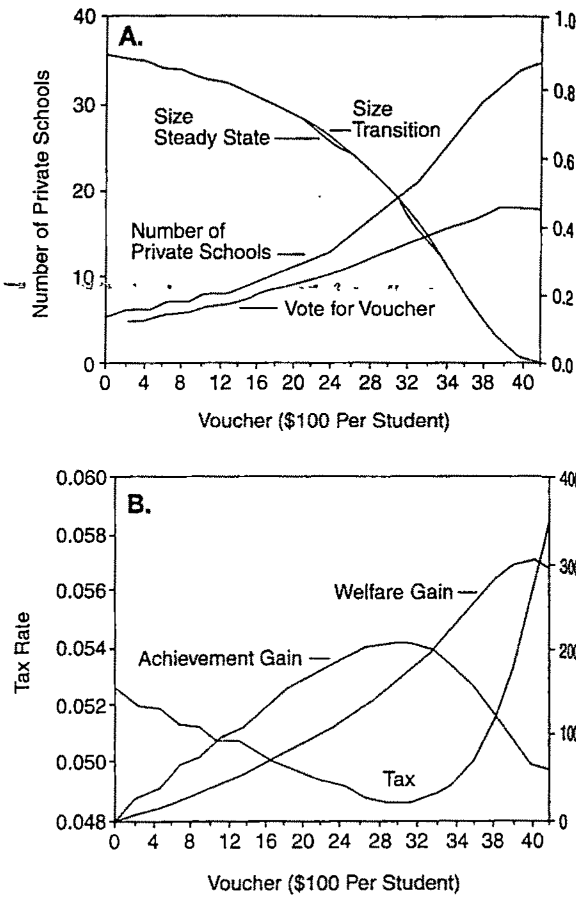


FIGURE 4. (A) PUBLIC-SECTOR SIZE IN TRANSITION AND STEADY STATE (RIGHT SCALE), VOTE FOR VOUCHER (RIGHT SCALE), AND NUMBER OF PRIVATE SCHOOLS (LEFT SCALE); (B) WELFARE AND ACHIEVEMENT GAIN (RIGHT SCALE) AND TAX RATE (LEFT SCALE) AS A FUNCTION OF VOUCHER SIZE

school with no voucher. A theme of our computational findings is that, while gains from vouchers result on average, there is a majority with relatively small loses and a minority with relatively large gains.

The bottom panel of Figure 4 also shows the per-student welfare gain, compensating variation plus net profit change (the latter relatively small), associated with the introduction of the voucher. The welfare gain associated with the maximum voucher is a relatively modest 0.5 percent of mean income, but as we will show, the distributional effects are more substantial. Welfare rises with the voucher until about 86 percent of the population attends school in the private sector. Apparently, negative entry externalities come to dominate welfare gains beyond this point (see footnote 21), as is manifest in rising average costs of private schools. The bottom panel of Figure 4 shows the tax rate at each voucher level. For low voucher levels, the cost of the voucher is more than offset by the reduction in public-school costs resulting from students who are induced by the voucher to choose private school, and the tax rate falls. Eventually, the tax rate must rise, however, since a sufficiently large voucher covers the cost of education for virtually all students in all schools.

Figure 4 also shows that per-student normed achievement, or equivalently, average future earnings of those who will work, rises and then declines with the voucher. The partition of students that maximizes earnings (gross or net of schooling costs) entails a strict hierarchy of schools with stratification by ability but not by income. This is implied by the complementarity of b and θ in our normed achievement function, and its property that peer effects impact per-student earnings more the higher is a school's θ.[28] Maximizing welfare differs from maximizing net earnings in our model because demand for student achievement depends on household income. By one interpretation, our model ascribes a household consumption motive to the educational achievement of the household's child(ren) (see footnote 8). The further partitioning of students caused by increasing the voucher will tend to increase welfare, as we have seen, but it has less-clear effects on ability stratification and, hence, on earnings. One effect of an increased voucher is to increase competition for high-ability students, which causes the boundary loci between incumbent private schools to become flatter (as further discussed below). This decreases the "degree" of ability stratification. Counteracting this is the migration of students into the private sector and the finer partition of students associated with entry of private schools. The latter effect dominates at low vouchers in our computations, and earnings rise initially with the voucher. The former effect comes to dominate for moderately large vouchers, and earnings decline with further increases in the voucher. Hence, while welfare is maximized at a voucher high enough that 86 percent of students attend private schools, earnings maximization necessitates a substantially lower voucher that draws only 47 percent of students to the private sector.

The aggregate effects on achievement and welfare summarized above are likely to be sensitive to the details of the specification of the achievement function. In our specification of normed achievement, gains from peer ability are proportionate to own ability. A specification of normed achievement in which high-ability students gain even more from high peer quality than do low-ability students would yield greater aggregate gains from sorting students by ability, for example. As we noted earlier, there is little empirical evidence on this issue.

Aggregate achievement is maximized by complete ability stratification in our Cobb-Douglas specification, as observed above. In our computational model, complete ability stratification increases (normed) aggregate achievement by 4.6 percent over complete mixing of students.[29] One-third of this potential gain is achieved in the public/private equilibrium without a voucher. As we illustrated in the bottom panel of Figure 4, aggregate achievement is nonmonotonic in the amount of the voucher.

[28] Proof and detailed analysis of these points is in an unpublished Appendix, which is available from the authors upon request. A nascent related literature studies the optimal grouping of workers of varying skills into firms (see Michael Kremer and Eric Maskin, 1995). Every student in a school has his own output (achievement), while workers in a firm have but one output. Hence, the results of the literature on worker grouping do not immediately map into the problem of student grouping.

[29] Normed aggregate achievement with complete mixing is given by

$$A_m \equiv K \int_0^\infty \exp\left[\frac{\gamma}{\beta}(\mu_b + \sigma_b^2/2)\right] b f_b(b)\, db$$

$$= K \exp\left[\left(1 + \frac{\gamma}{\beta}\right)(\mu_b + \sigma_b^2/2)\right]$$

where $f_b(b)$ is the lognormal density function. Normed aggregate achievement with complete ability stratification is given by

$$A_s \equiv K \int_0^\infty b^{(\gamma/\beta)+1} f_b(b)\, db$$

$$= K \exp\left[\left(\frac{\gamma}{\beta} + 1\right)\mu_b + \left(\frac{\gamma}{\beta} + 1\right)^2 \sigma_b^2/2\right].$$

Hence,

$$\frac{A_s}{A_m} = \exp\left[\frac{\gamma}{\beta}\left(\frac{\gamma}{\beta} + 1\right)\frac{\sigma_b^2}{2}\right] \approx 1.046$$

for our calibration. In steady state with complete mixing, one can show that σ_b^2 is invariant to γ/β. Hence, the preceding formula can be used to determine how potential gains to ability stratification vary with γ/β. Such gains increase with γ/β (e.g., to about 45 percent for $\gamma = \beta$).

Maximum achievement occurs at a voucher of $2,800, and this maximum is 50 percent of the potential achievement gain of moving from complete mixing to complete ability stratification. We emphasize that such aggregate gains can undoubtedly be made either larger or smaller by varying the relative benefit of peer quality to students of different ability levels.

The remaining figures illustrate the distributional effects of the voucher. The effects of the voucher on achievement and welfare can be divided into the impacts on the four groups illustrated in the upper panel of Figure 5. Area A contains students who are in the public schools before and after the introduction of a $2,000 voucher. Areas B and C combined are students who switch from public to private school when the voucher is introduced. Area D contains students in the private-school sector before and after the introduction of the voucher.

The primary gains in achievement accrue to students who switch from the public- to the private-school sector with introduction of the voucher (areas B and C). They have achievement gains ranging between 12.9 percent and 20.2 percent. The major losses in achievement are experienced by students who remain in the public school after the voucher is introduced. They all experience a 4.9-percent loss. The latter group is, of course, much larger than the former.

The welfare effects (measured by compensating variation) are distributed somewhat differently. Students who remain in the public sector (area A) all experience welfare losses. The quality of the school they attend has deteriorated, and tax changes are small as discussed above. Since public schools charge a price of zero, the voucher does not reduce the cost of education for public-school students.

Paradoxically, some of those who switch from the public to the private sector (area B) are also made worse off. Their alternatives are adversely affected by the voucher. They can either stay in a public-school system of reduced quality, or they can pay tuition at private school. They choose the latter, but the voucher defrays only a portion of the cost. Thus, while they have large achievement gains, those gains are more than offset by the reduction in income net of tuition.[30] For each income level the largest loss within this group is sustained by the students at the lower boundary of area B. By boundary indifference, the magnitude of loss for a student on this lower boundary is the same as for a comparable student on the upper boundary of region A. Students on the boundary between regions B and C neither gain nor lose from the voucher.

The remaining two groups (C and D) gain from the voucher. The largest gains as a proportion of income accrue to high-ability, low-income households. As the voucher increases the demand for private education, it increases competition for high-ability students and the financial aid they receive.[31] The highest-ability students experience a tuition reduction that is almost twice the amount of the voucher. The greatest gains are thus in the lower-right portion of the upper panel of Figure 5.

The bottom panel of Figure 5 illustrates further the distributional effects of the $2,000 voucher. Compensating variation as a percentage of income is plotted for four different income levels as a function of ability. This figure demonstrates that the gains accrue to those with high ability, and among high-ability households, the gains are proportionately greater for low-income households. Losses are realized by households that remain in the public-school sector due to the decline in public-school peer quality induced by the voucher. Again, those in area B in the upper panel are also losers. The proportionate loss for the latter two sets of households is low, but they make up a majority of the population.

[30] This result is similar to Benabou's (1996a) finding that equilibrium segregation across communities (and schools) by endowed human capital (ability) can cost high-human-capital types more in housing-price premia than they gain relative to an allocation without segregation.

[31] For example, when the voucher is increased from 0 to $1,800, the η's in the top four schools rise from 360, 310, 282, and 261 (dollar discount per unit of ability; see Figure 2) to 398, 351, 326, and 308, respectively. Interestingly, the increased competition for high-ability students actually reduces the quality of the top schools, as entrants bid some of these students away, and the boundary loci between private schools become flatter. Compare the θ's of the top schools in the two panels of Figure 5. This phenomenon persists in the steady state.

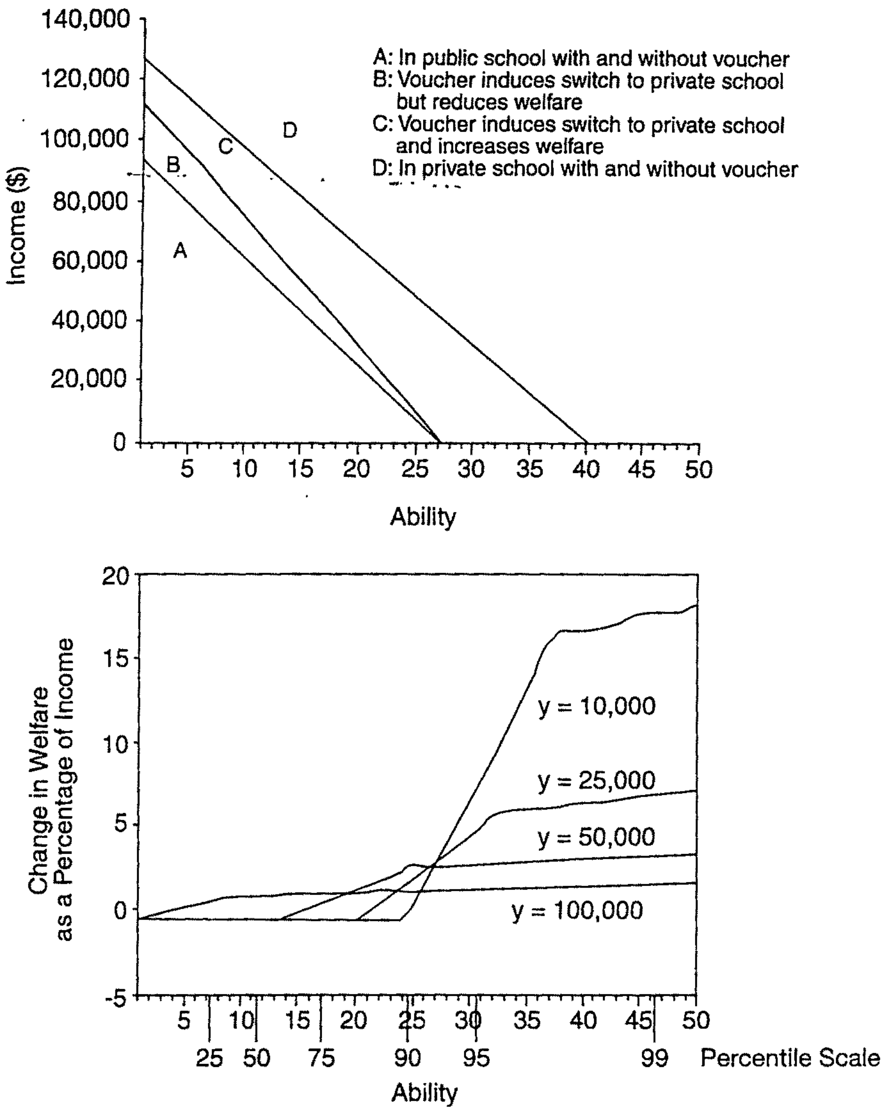


FIGURE 5. BOUNDARY LOCI BETWEEN PUBLIC AND PRIVATE SCHOOLS WITHOUT AND WITH A $2,000 VOUCHER AND LOCUS OF HOUSEHOLDS INDIFFERENT TO PROVISION OF THE VOUCHER (UPPER PANEL) AND CHANGE IN WELFARE AS A FUNCTION OF ABILITY FOR FOUR DIFFERENT INCOME LEVELS (LOWER PANEL)

Figure 6 illustrates the distributional effects of the voucher in steady state. For each ability and a voucher level, we calculate the expected steady-state income of a student as follows. Given a student ability and knowing the steady-state distribution of parental income, we calculate the probability that a student will attend each of the available schools. Using this, each school's quality, and the student's ability, we calculate expected income. Figure 6 shows the percentage change in expected steady-state income relative to the zero-voucher steady state. One would expect gains to accrue to the bulk of the relatively highest-ability students since the relatively highest-ability students are most likely to attend a higher-quality school as a result of the introduction of the voucher. Lower-ability students comprising approximately 70 percent of the population are made worse off because they are likely either to remain in the public sector when the voucher

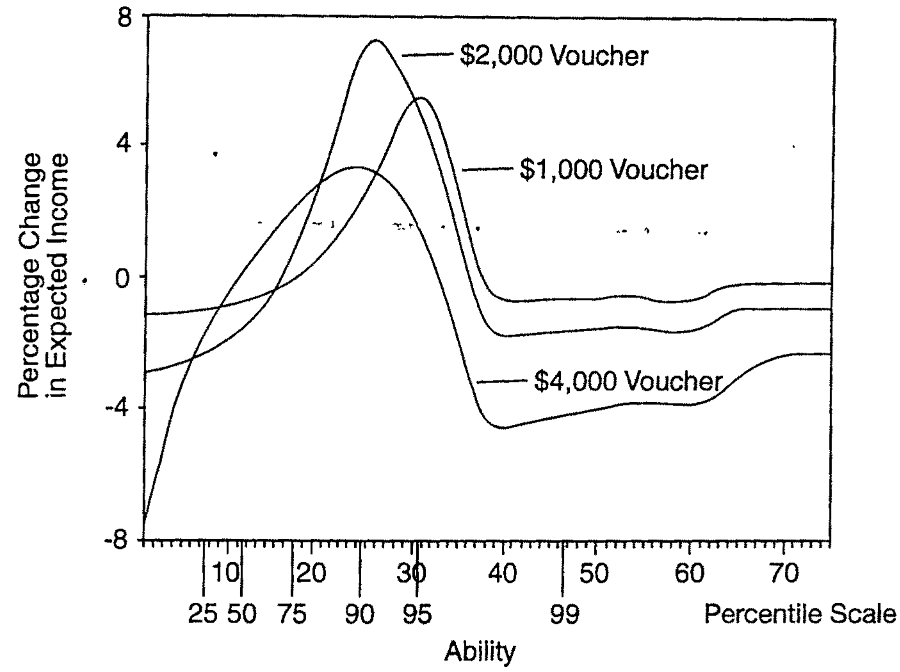


FIGURE 6. PERCENTAGE STEADY-STATE CHANGE IN EXPECTED INCOME FROM INTRODUCING VOUCHERS FOR STUDENTS OF DIFFERING ABILITY LEVELS

is introduced (a public sector of diminished quality) or to enter a low-quality private school. The top 2–3 percent of the ability distribution (abilities 35 and higher) also have lower expected income because the very top schools that they will attend decline somewhat in quality (see footnote 30).

IV. Concluding Remarks

A recent Department of Education study reports that nearly half of all adult Americans read and write so poorly that they are unable to function effectively in the workplace, and not surprisingly, that many of them live in poverty (see *Newsweek*, 1993; *Wall Street Journal*, 1993). This depressing statistic, like many others before it, has led to calls for change in the U.S. educational system. There is no shortage of reform proposals. The research challenge is to develop models that provide systematic links connecting preferences, the educational "production" process, costs, and institutional structure to consequences. Such models should develop testable predictions for validating or rejecting the elements of beliefs that give rise to policy proposals, and they should facilitate systematic comparison and evaluation of policy alternatives.

One goal of our research is to provide a useful foundation both in permitting formal analysis of educational policy issues and in guiding empirical work. We are introducing inputs into the model so that schools can compete for students by varying inputs as well as tuition policy. We (Epple and Romano, 1995) are also enriching the model to contrast equilibrium in an open-enrollment system (as in the model in this paper) to a neighborhood system in which students may only attend the school in the geographic neighborhood in which they reside. With Elizabeth Newlon, we (Epple et al., 1997) have analyzed equilibrium when public schools adopt tracking by ability.[32] These extensions are discussed further below. The model can be extended to allow for multiple peer characteristics. This is likely to lead to an equilibrium with a more diverse set of private schools, diversity seen by pro-

[32] See Gamoran (1992) for an interesting empirical analysis of tracking.

ponents of vouchers as a potential benefit. A further extension lets teachers vary in skill with student achievement increasing in both the school's teacher–student ratio and the mean skill of teachers. If a teacher's utility depends on the abilities of students taught, private schools are advantaged in hiring highly skilled teachers. Allowing a teacher's utility to depend also on the skills of colleagues is a key step toward applying the model to higher education. In analyzing higher education, we also intend to introduce a quasi-public sector (state schools) subsidized by tax dollars but facing egalitarian dictates that constrain admission and tuition policies.

Regarding implications for empirical work, our analysis delivers several predictions summarized in the propositions we have presented. These include a negative correlation between income and ability within each private school, a negative correlation between private-school size and quality, tuitions declining in ability in all private schools, tuitions increasing in income for the best private school in each geographic market, and little dependence of tuition on income in the remaining private schools in each market, a scale of operation in all private schools lower than that which minimizes average cost, a strict hierarchy of schools, stratification by income, and stratification by ability.[33]

As one of our referees stated, what we would like to know about a voucher program is "who gains, who loses, and how does it add up?" Earlier, we discussed aggregate effects in detail. Here we discuss further the redistributional effects. In this paper, we consider an open-enrollment public system in which public and private schools are equally effective in delivering education. Our model implies that a voucher program will result in entry of new private schools and movement of students from the public to the private sector. Students remaining in the public sector are those with relatively low income and low ability, and those students experience losses. Because vouchers increase the premium on ability, the greatest proportionate gains from the voucher accrue to low-income, high-ability students.

How sensitive are our results likely to be to the assumption of a homogeneous open-enrollment public-school system? The conclusion that low-income, high-ability students experience large proportionate gains is likely to hold regardless of the organization of the public-school system. However, the distribution of losses from a voucher program is likely to depend on the organization of public schools. For example, most public schools in the United States place students of differing abilities into different "tracks." Applying the framework developed here to study tracking reveals the following (Epple et al., 1997). The largest proportionate losses from a voucher program are to students who remain in the high-ability public-school track when the voucher is introduced, because the voucher draws the most-able students from the high track into the private sector. In the presence of tracking, the voucher has very little impact on the low-income, low-ability students. Public schools consign these students to a low-ability peer group, and the voucher leads to modest decline in the quality of that peer group. The loss to such students from a voucher program is much smaller than the losses they experience from introduction of public-school tracking.

Consider now a neighborhood public-school system, without tracking, that assigns students to a school based on the neighborhood in which they reside. We have shown (Epple and Romano, 1995) that, even absent differences in expenditure across schools, Tiebout sorting will create a public-school hierarchy deriving from peer-group differences when income and student ability are positively correlated. We anticipate that the introduction of a significant private sector, as supported by vouchers, would again benefit most poor, high-ability students. How poor, low-ability students would be affected relative to the case of open-enrollment public schools is less clear. As with tracking, the latter students begin with a weaker peer group. Thus, unfortunately, those who remain in public school when vouchers covering part of educational costs are introduced may have

[33] The specific properties of the tuition functions are developed in Epple and Romano (1993).

little to lose. On the other hand, their most-able peers will be the first to enter the private sector. This could imply even further losses to those who remain behind. We think this is an important issue for research, since neighborhood school systems continue to dominate in the United States.

It is often argued that large central-city public-school systems in the United States are ineffective in delivering education, and that the failings of these schools are visited primarily on disadvantaged students who reside in central cities. Clearly, if vouchers lead private schools to supplant ineffective public schools or inspire better performance from such ineffective schools, then a voucher program would lead to widely distributed gains. In the presence of public-school tracking, for example, such gains, even if relatively modest, might be sufficient to offset the effects of diminished peer quality experienced by low-income, low-ability students. Such gains would need to be more substantial, however, to offset the losses experienced by students who remain in the high-ability public-school track after vouchers are introduced.

Whether such gains in technical efficiency from voucher programs would be substantial is an open question (see footnote 4). Our paper stresses the allocative effects of vouchers and shows that vouchers could have significant distributional consequences. It indicates a need for continued effort to quantify the effects of private-sector competition in education and to quantify the effects of school quality on students of differing ability levels. There are also implications for the design of voucher systems intended to promote alternative goals. If vouchers are intended to improve technical efficiency without increasing ability segregation, then less-able students will need more financial assistance (or equivalent controls must be enacted). If ability segregation is to be further promoted to increase aggregate achievement, then vouchers will need to be income dependent. Hopefully, our framework will facilitate the investigation of such voucher-design issues, as well as the investigation of other proposals affecting the entry and exit of schools or student access to schools.

Appendix

PROOF OF PROPOSITION 1:

We show first that the public school has the worst peer group. Suppose to the contrary that

$$\theta_0 \geq \theta_1 \in \underset{i \in \{1,2,\ldots,n\}}{\text{MIN}} \theta_i .$$

Any student who would have to pay a positive price to attend private school 1 would prefer the free public school. Hence, private school 1 could not be profitable.

Showing that a strict hierarchy of private schools characterizes equilibrium is considerably more involved. The proof is by contradiction, so assume that $\theta_i = \theta_j$ for some $i \neq j$, $i, j = 1, 2, \ldots, n$. We show that this implies

$$\text{(A1)} \quad p_i(b, y) = p_j(b, y) = \text{MC}_i(b) = \text{MC}_j(b) \ \forall \ (b, y)$$

with $\alpha_i(b, y) > 0$ and/or $\alpha_j(b, y) > 0$.

We will go on to show by construction that (A1) implies that school i (or j) can increase profits by admitting and expelling certain students, contradicting profit maximization.

Condition (6a) implies $p_i(b, y) = p_j(b, y) \ \forall \ (b, y)$ since $\theta_i = \theta_j$. Condition (6b) and market clearance imply that $\text{MC}_i(b) \geq p_i(b, y) = p_j(b, y) \geq \text{MC}_j(b)$ for students who attend school j and, analogously, $\text{MC}_j(b) \geq p_i(b, y) = p_j(b, y) \geq \text{MC}_i(b)$ for students who attend school i. The linearity of MC(b) implies: (a) $\text{MC}_j(b) = \text{MC}_i(b) \ \forall \ b$; (b) $\text{MC}_i(b) < \text{MC}_j(b) \ \forall \ b$ (or equivalently, the reverse); or (c) $\text{MC}_i(b) > (=) (<) \text{MC}_j(b)$ as $b > (=) (<) b'$, for some b' (again, reversing i and j provides an equivalent). Case (b) precludes school j from admitting any students and can be rejected. Case (c) precludes $\theta_i = \theta_j$, since school i would admit only student types with $b \leq b'$, and school j would admit only student types with $b \geq b'$. This leaves case (a), which then implies (A1).

We now show that student types (b_1, y_1) and (b_2, y_2) exist, with $\alpha_j(b_1, y_1) \in (0, 1]$ and $\alpha_i(b_2, y_2) \in (0, 1]$, such that school i can increase profits by admitting the same number of (b_1, y_1) types as it expels of (b_2, y_2) types. School i will increase profits then by admitting some students who initially attend j but expelling the same number of its own students. Note that, since tuitions equal effective marginal costs in school i for both such types of students, it is a local nonconcavity of profits that must drive this result. Hence, we will consider higher-order derivatives.

Proceeding, we have

$$\text{(A2)}\quad \frac{\partial \pi_i}{\partial[\alpha_i(b_1, y_1) f(b_1, y_1)]} = p_i^*(b_1, y_1, \theta_1) - V'(k_i) - (\theta_i - b_1)\frac{1}{k_i}\iint_S \frac{\partial p_i^*}{\partial \theta_i}\alpha_i f\, db\, dy.$$

Denote the expression (A2) with π_1^i [the subscript since the first variation is for type (b_1, y_1)]; π_2^i is analogous. Using (5c) and (5d), one obtains

$$\text{(A3)}\quad \pi_{11}^i \equiv \frac{\partial^2 \pi_i}{\partial^2[\alpha_i(b_1, y_1) f(b_1, y_1)]} = -V''(k_i) + 2\left(\frac{b_1 - \theta_i}{k_i}\right) \times \left[\frac{\partial p_i^*(b_1, y_1, \theta_i)}{\partial \theta_i} - \frac{1}{k_i}\iint_S \frac{\partial p_i^*}{\partial \theta_i}\alpha_i f\, db\, dy\right] + \left[\left(\frac{b_1 - \theta_i}{k_i}\right)^2 \times \iint_S \frac{\partial^2 p_i^*}{\partial \theta_i^2}\alpha_i f\, db\, dy\right]$$

$$\text{(A4)}\quad \pi_{12}^i \equiv \frac{\partial^2 \pi_i}{\partial[\alpha_i(b_1, y_1) f(b_1, y_1)]\,\partial[\alpha_i(b_2, y_2) f(b_2, y_2)]} = -V''(k_i) + \frac{\partial p_i^*(y_1, b_1, \theta_i)}{\partial \theta_i}\left(\frac{b_2 - \theta_i}{k_i}\right) + \frac{\partial p_i(y_2, b_2, \theta_i)}{\partial \theta_i}\left(\frac{b_1 - \theta_i}{k_i}\right) - \left(\frac{b_2 + b_1 - 2\theta_i}{k_i}\right)\frac{1}{k_i}\iint_S \frac{\partial p_i^*}{\partial \theta_i}\alpha_i f\, db\, dy + \frac{(b_1 - \theta_i)(b_2 - \theta_i)}{k_i^2}\iint_S \frac{\partial^2 p_i^*}{\partial \theta_i^2}\alpha_i f\, db\, dy.$$

The expression for π_{22}^i is analogous to (A3). Let Δ_i equal the change in the number (density) of types (b_i, y_i) enrolled in school i. For Δ_i's sufficiently small, Taylor's theorem implies the sign of the change in school i's profit, $\Delta\pi_i$, will be the same as the sign of

$$\text{(A5)}\quad \pi_1^i \Delta_1 + \pi_2^i \Delta_2 + \tfrac{1}{2}[\pi_{11}^i(\Delta_1)^2 + \pi_{22}^i(\Delta_2)^2 + 2\pi_{12}^i \Delta_1 \Delta_2].$$

We know that $\pi_1^i = \pi_2^i \doteq 0$ since prices equal effective marginal costs. We will consider admission changes such that $\Delta_1 = -\Delta_2$. Hence, (A5) simplifies to $\frac{1}{2}(\Delta_1)^2[\pi_{11}^i + \pi_{22}^i - 2\pi_{12}^i]$. Substituting and simplifying yields

$$\text{(A6)}\quad \operatorname{sign}\Delta\pi_i = \operatorname{sign}\left\{\frac{2}{k_i}(b_1 - b_2)\left[\frac{\partial p_i^*(y_1, b_1, \theta_i)}{\partial \theta_i} - \frac{\partial p_i^*(y_2, b_2, \theta_i)}{\partial \theta_i}\right] + \left(\frac{b_1 - b_2}{k_i}\right)^2 \iint_S \frac{\partial^2 p_i^*}{\partial \theta_i^2}\alpha_i f\, db\, dy\right\}.$$

Suppose that y_1 is greater than y_2 but very close to it, and that b_1 is greater than b_2 but closer still to b_2, formally, of lower-order

difference. In a moment, we show that school i can substitute students this way. From (6a),

$$\frac{\partial p_i^*}{\partial \theta_i} = \frac{\partial U/\partial \theta_i}{\partial U/\partial y_t}.$$

Lemma 5 in the Appendix to Epple and Romano (1993) shows that $p_i^*(b, y, \theta)$ is continuous in (b, y). By SCI and this lemma (keeping in mind that b_1 is extremely close to b_2), the first term in (A6) is positive. Moreover, it will dominate the second term due to the lower order of the difference $(b_1 - b_2)$ and, again, using the lemma.

The result then follows so long as such a substitution of students by school i is feasible. Such a substitution is clearly feasible if there is a positive measure of type space over which both schools i and j admit students. Suppose, however, that no such overlap exits. School i accomplishes an equivalent substitution of students in what can be thought of as two steps. It first expels half of all of its students and admits half of all of the students who attend school j initially. Since $\theta_i = \theta_j$ and $\mathrm{MC}_i = \mathrm{MC}_j$, this step has no effect on i's profits or θ_i, k_i, and η_i (hence MC_i). Now the profit-increasing substitution is feasible, which constitutes the second step. In this case, the variation that increases profits is just more involved.

PROOF OF PART (ii) OF PROPOSITION 3:

Household choice is as though tuitions equal effective marginal cost by part (iii) of Proposition 2; that is, type (b, y) chooses school i over j if and only if $U(y_t - \mathrm{MC}_i(b), a(\theta_i, b)) \geq U(y_t - \mathrm{MC}_j(b), a(\theta_j, b))$. Keep in mind our convention that $\mathrm{MC}_0(b) \equiv 0\ \forall\, b$ and let $\eta_0 \equiv 0$. Also, let $p_i^{\mathrm{d}} \equiv V'(k_i) + \theta_i\eta_i$, $i = 1, \ldots, n$, and $p_0^{\mathrm{d}} \equiv 0$, so that the first argument in the utility function is $y_t + \eta b - p^{\mathrm{d}}$. Choice over schools by type (b, y) facing tuition $\mathrm{MC}_i(b)$ is equivalent to choice over schools by type $(b, y_t + \eta_i b)$ facing *type-independent* tuitions p_i^{d}. We map the school-choice problem into (θ, p^{d})-space and show that a standard single-crossing argument implies SBA.

Since $0 = \eta_0 < \eta_1 \leq \eta_2 \leq \cdots \leq \eta_n$, an infinite number of nondecreasing and differentiable functions $\hat{\eta}(\theta)$ exist satisfying $\hat{\eta}(\theta_i) = \eta_i$, $i = 0, 1, \ldots, n$. Choose any such function. Choice of schools is equivalent to

$$\text{(A7)}\quad \max_{i \in \{0,1,\ldots,n\}} \left[U(y_t + \hat{\eta}(\theta_i)b - p_i^{\mathrm{d}}, a(\theta_i, b))\right].$$

Indifference curves in (θ, p^{d})-space are upward sloping:

$$\text{(A8)}\quad \left.\frac{dp^{\mathrm{d}}}{d\theta}\right|_{U(\cdot)=\bar{U}} = b\hat{\eta}' + \frac{(\partial U(\cdot)/\partial a)(\partial a/\partial \theta)}{\partial U(\cdot)/\partial y_t}$$

where $U(\cdot)$ has the same arguments as in (A7). The first term in (A8) is nonnegative since $\hat{\eta}(\theta)$ is nondecreasing, and the second term is obviously positive. These indifference curves also have slopes that are strictly increasing in b for $\theta > \theta_0$ and weakly increasing at $\theta = \theta_0$. Differentiation of (A8) with respect to b yields:

$$\text{(A9)}\quad \hat{\eta}\frac{\partial}{\partial y_t}\left[\frac{(\partial U(\cdot)/\partial a)(\partial a/\partial \theta)}{\partial U(\cdot)/\partial y_t}\right] + \hat{\eta}' + \frac{\partial}{\partial b}\left[\frac{(\partial U(\cdot)/\partial a)(\partial a/\partial \theta)}{\partial U(\cdot)/\partial y_t}\right]$$

where b in the first argument of $U(\cdot)$ is held constant in the latter term. The first term in (A9) is strictly (weakly) positive by SCI for $\theta > (=)\ \theta_0$, and the second two terms are weakly positive by $\hat{\eta}' \geq 0$ and W-SCB, respectively. Given that slopes of indifference curves increase with b, SBA follows by a standard single-crossing argument.

Discussion of Existence.—Here we describe more fully the existence problem, potential resolutions of it, and the related literature on club theory. Consider a quasi-equilibrium satisfying the exact-equilibrium requirements UM, PSP, and MC, but substituting an exogenously determined number of private schools for ZΠ and replacing ΠM with a requirement

of local profit maximization [i.e., (6a)–(6c)] for all private schools. Proposition 1 would continue to apply, implying that the private schools serve different niches of students. Unless the distribution of types is "just so," the profitability of serving the different niches will vary. Each private school is maximizing profits locally, but the profit function (which the utility-taking schools share) has multiple peaks (in the space of functions $[p(b,y), \alpha(b,y)]$), and only one school is at the global maximum. The other private schools are at lower-profit peaks and would have an incentive to mimic the tuition/admission policy of the highest-profit school. This problem persists regardless of the (finite) number of schools, so that $Z\Pi$ cannot generally be satisfied.

The existence problem is fundamental to club economies (see below), and no easy resolution presents itself. The epsilon equilibrium described in the text keeps private schools at their local maxima and, roughly, allows entry until the profit peaks and their differences are minimized. One could make ε arbitrary low by reducing the fixed costs, but this presents computational difficulties and sacrifices realism. Related to this, one might also assume constant costs of schooling. This would lead to an infinite number of schools serving infinitely refined peer groups. We have made some progress on such a specification in a special case with the help of an anonymous referee. We believe exact equilibrium exists in such specifications, but this has not been proved. While this model is extremely interesting, it is quite complex and not yet tractable.

The natural game-theoretic specification also encounters existence problems. Suppose that, in two stages, private schools first commit to enter (e.g., pay fixed costs) and then play a Nash game in tuition functions. We believe that the second-stage profit functions will continue to be multipeaked, and Bertrand-like lack of pure-strategy equilibrium will arise. Scotchmer (1985) shows existence of equilibrium in pure strategies in the case of clubs with anonymous crowding using a game-theoretic approach. The difference is that incumbent clubs earn the same level of profits, a property that our model will not generally have, because the variety of types induces differentiated "clubs."

The problem we encounter is familiar to club economies with *nonanonymous* crowding. Either economies of scale or indivisibility of club members generally leads to existence problems in such economies, both causes detailed in Scotchmer and Wooders (1987) and Scotchmer (1997). Our continuum of agent types also implies a continuum of each type, eliminating the latter existence problem as in Scotchmer (1986). But the first existence problem remains. In lieu of exact equilibria, epsilon-competitive equilibria are also studied in this literature, with the difference that it is club members rather than club owners who do not fully optimize. A correspondence between exact equilibrium and the exact core and between epsilon equilibrium and an epsilon core is demonstrated in this literature. This normative focus highlights consumer behavior rather than "firm" behavior, and consumers are the natural candidates for deviation from full optimization. Our positive focus provides a more salient role for firms (private-school owners), and they are more convenient candidates for deviation than the infinity of households. Hence, our epsilon equilibrium has households fully optimize but not private schools.

REFERENCES

Arnott, Richard and Rowse, John. "Peer Group Effects and Educational Attainment." *Journal of Public Economics*, April 1987, *32*(3), pp. 287–305.

Baker, Steven. "Reading, Writing, and Rancor." *Business Week*, September 18, 1995, p. 54.

Baum, Sandra R. and Schwartz, Saul. "Merit Aid to College Students." *Economics of Education Review*, 1988, *7*(1), pp. 127–34.

Benabou, Roland. "Equity and Efficiency in Human Capital Investment: The Local Connection." *Review of Economic Studies*, April 1996a, *63*(2), pp. 237–64.

———. "Heterogeneity, Stratification, and Growth: Macroeconomic Implications of Community Structure and School Finance." *American Economic Review*, June 1996b, *86*(3), pp. 584–609.

Bergstrom, Theodore; Rubinfeld, Daniel L. and Shapiro, Perry. "Micro-Based Estimates of Demand Functions for Local School

Expenditures." *Econometrica*, September 1982, *50*(5), pp. 1183–1205.

Chubb, John E. and Moe, Terry M. *Politics, markets, and America's schools*. Washington, DC: Brookings Institution, 1990.

Clotfelter, Charles T. "Financial Aid and Public Policy," in Charles T. Clotfelter, Ronald G. Ehrenberg, Malcolm Getz, and John J. Siegfried, eds., *Economic challenges in higher education*. Chicago: University of Chicago Press, 1991, pp. 89–123.

Coons, John E. and Sugarman, Stephen D. *Education by choice: The case for family control*. Berkeley, CA: Institute of Governmental Studies Press, 1978.

de Bartolomé, Charles A. M. "Equilibrium and Inefficiency in a Community Model with Peer Group Effects." *Journal of Political Economy*, February 1990, *98*(1), pp. 110–33.

De Witt, Karen. "Fanfare and Catcalls for Voucher Plan." *New York Times*, June 24, 1992, p. A17.

Eden, Ben. "How to Subsidize Education: An Analysis of Voucher Systems." Working paper, University of Iowa, 1992.

Epple, Dennis; Newlon, Elizabeth and Romano, Richard E. "Ability Tracking, School Competition, and the Distribution of Educational Benefits." Working paper, Carnegie Mellon University, February 1997.

Epple, Dennis and Romano, Richard E. "Competition Between Private and Public Schools, Vouchers, and Peer Group Effects." Working paper, University of Florida, September 1993.

______. "Public School Choice and Finance Policies, Neighborhood Formation, and the Distribution of Educational Benefits." Working paper, University of Florida, July 1995.

______. "Ends Against the Middle: Determining Public Service Provision When There Are Private Alternatives." *Journal of Public Economics*, November 1996, *62*(3), pp. 297–325.

Evans, William N.; Oates, Wallace E. and Schwab, Robert M. "Measuring Peer Group Effects: A Study of Teenage Behavior." *Journal of Political Economy*, October 1992, *100*(5), pp. 966–91.

Evans, William N. and Schwab, Robert M. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *Quarterly Journal of Economics*, November 1995, *110*(4), pp. 941–74.

Fernandez, Raquel and Rogerson, Richard. "Income Distribution, Communities, and the Quality of Public Education: A Policy Analysis." *Quarterly Journal of Economics*, February 1996, *111*(1), pp. 135–64.

Figlio, David N. and Stone, Joe A. "School Choice and Student Performance: Are Private Schools Really Better?" Working paper, University of Washington, 1997.

Forbes, Steve. "Good News from Jersey City." *Forbes*, March 28, 1993, *153*(7), p. 28.

Friedman, Milton. *Capitalism and freedom*. Chicago: University of Chicago Press, 1962.

Fujita, Masahisa. *Urban economic theory: Land use and city size*. Cambridge, MA: Cambridge University Press, 1989.

Gamoran, Adam. "The Variable Effects of High School Tracking." *American Sociological Review*, December 1992, *57*(6), pp. 812–28.

Gamoran, Adam and Berends, Mark. "The Effects of Stratification in Secondary Schools: Synthesis of Survey and Ethnographic Research." *Reviews of Educational Research*, Winter 1987, *57*(4), pp. 415–535.

Glazerman, Steven and Meyer, Robert H. "Public School Choice in Minneapolis." Working paper, University of Chicago, October 1994.

Glomm, Gerhard and Ravikumar, B. "Opting Out of Publicly Provided Services: A Majority Voting Result." *Social Choice and Welfare*, 1998 (forthcoming).

Green, Jay P.; Peterson, Paul E.; Du, Jiangtao; Boeger, Leesa and Frazier, Curtis L. "The Effectiveness of School Choice in Milwaukee: A Secondary Analysis of Data from the Program's Evaluation." Unpublished manuscript prepared for August–September 1996 meetings of American Political Science Association, San Francisco, CA, August 1996.

Hanushek, Eric. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, September 1986, *24*(3), pp. 1141–77.

Henderson, Vernon; Mieszkowski, Peter and Sauvageau, Yvon. "Peer Group Effects and Educational Production Functions." *Jo*

nal of Public Economics, August 1978, *10*(1), pp. 97–106.
Hoxby, Caroline M. "Do Private Schools Provide Competition for Public Schools?" National Bureau of Economic Research (Cambridge, MA) Working Paper No. 4978, 1994.
——. "The Effects of Private School Vouchers on Schools and Students," in Helen Ladd, ed., *Holding schools accountable*. Washington, DC: Brookings Institution, 1996, pp. 177–208.
Ireland, Norman J. "The Mix of Social and Private Provision of Goods and Services." *Journal of Public Economics*, November 1990, *43*(2), pp. 201–19.
Johnson, Daphne. *Parental choice in education*. London: Unwin Hyman, 1990.
Kach, Nick and Mazurek, Kas. "Canada: A Social and Educational Profile," in N. Kach, K. Mazurek, Robert S. Patterson, and Ivan Defaveri, eds., *Essays on Canadian education*. Calgary, Alberta: Detselig, 1986, pp. 1–22.
Kane, Thomas J. "Comments on Chapters Five and Six," in Helen Ladd, ed., *Holding schools accountable*. Washington, DC: Brookings Institution, 1996, pp. 209–20.
Kenny, Lawrence. "Economies of Scale in Schooling." *Economics of Education Review*, Winter 1982, *2*(1), pp. 1–24.
Koretz, Daniel M. *Educational achievement: Explanations and implications of recent trends*. Washington, DC: Congressional Budget Office, Congress of the United States, August 1987.
Kremer, Michael and Maskin, Eric. "Segregation by Skill and the Rise in Inequality." Working paper, Harvard University, December 1995.
Kulik, Chen-Lin and Kulik, James A. "Effects of Ability Grouping on Secondary School Students: A Meta-analysis of Evaluation Findings." *American Education Research Journal*, Fall 1992, *19*(3), pp. 415–28.
Kulik, James A. and Kulik, Chen-Lin. "Effects of Accelerated Instruction on Students." *Review of Educational Research*, Fall 1984, *54*(3), pp. 409–25.
Lieberman, Myron. *Privatization and educational choice*. New York: St. Martin's, 1989.
Manski, Charles F. "Educational Choice (Vouchers) and Social Mobility." *Economics of Education Review*, December 1992, *11*(4), pp. 351–69.
Moreland, Richard L. and Levine, John M. "The Composition of Small Groups," in E. J. Lawler, B. Markovsky, C. Ridgeway, and H. Walker, eds., *Advances in group processes*, Vol. 9. Greenwich, CT: JAI Press, 1992, pp. 237–80.
Nathan, Joe. "Chartered Public Schools: A Brief History and Preliminary Lessons." Working paper, University of Minnesota, October 1994.
National Commission on Excellence in Education. *A nation at risk: The imperative for educational reform*. A Report to the Nation and the Secretary of Education." Washington, DC: U.S. Department of Education, 1983.
Newsweek. "Dumber Than We Thought." *Newsweek*, September 20, 1993, *122*(12), pp. 44–45.
——. "Taking Public Schools Private." *Newsweek*, June 20, 1994, *123*(25), p. 77.
Oakes, Jeannie. "Tracking in Secondary Schools: A Contextual Perspective." *Educational Psychologist*, April 1987, *22*(2), pp. 129–53.
Rothschild, Michael and White, Lawrence J. "The Analytics of the Pricing of Higher Education and Other Services in Which the Customers Are Inputs." *Journal of Political Economy*, June 1995, *103*(3), pp. 573–623.
Rubinfeld, Daniel and Shapiro, Perry. "Micro-Estimation of the Demand for Schooling: Evidence from Michigan and Massachusetts." *Regional Science and Urban Economics*, August 1989, *19*(3), pp. 381–98.
Schwartz, J. Brad. "Wealth Neutrality in Higher Education: The Effects of Student Grants." *Economics of Education Review*, 1986, *5*(2), pp. 107–17.
Scotchmer, Suzanne. "Profit-Maximizing Clubs." *Journal of Public Economics*, June 1985, *27*(1), pp. 25–45.
——. "Nonanonymous Crowding: The Core With A Continuum of Agents." Harvard Institute of Economic Research Discussion Paper No. 1236, 1986.
——. "Public Goods and the Invisible Hand," in J. Quigley and E. Smolensky,

eds., *Modern public finance*. Cambridge, MA: Harvard University Press, 1994, pp. 98–119.

______. "On Price-Taking Equilibria in Club Economies with Nonanonymous Crowding." *Journal of Public Economics*, June 1997, *65*(1), pp. 75–87.

Scotchmer, Suzanne and Wooders, Myrna H. "Competitive Equilibrium and the Core in Club Economies with Nonanonymous Crowding." *Journal of Public Economics*, November 1987, *34*(2), pp. 159–73.

Slavin, Robert E. "Ability Grouping and Student Achievement in Elementary Schools: A Best-Evidence Synthesis." *Review of Educational Research*, Fall 1987, *57*(3), pp. 293–336.

______. "Achievement Effects of Ability Grouping in Secondary Schools: A Best-Evidence Synthesis." *Review of Education Research*, Fall 1990, *60*(3), pp. 471–99.

Solon, Gary. "Intergenerational Income Mobility in the United States." *American Economic Review*, June 1992, *82*(3), pp. 393–409.

Sorensen, Aage B. "The Organizational Differentiation of Students in Schools," in P. van den Eeden and H. Oosthoek, eds., *Education from the multilevel perspective: Models, methodology, and empirical findings*. London: Gordon and Breach, 1984, pp. 25–43.

Sorensen, Aage B. and Hallinan, Maureen T. "Effects of Ability Grouping on Growth in Academic Achievement." *American Educational Research Journal*, Winter 1986, *23*(4), pp. 519–42.

Statistical Abstract of the United States: 1991. Washington, DC: U.S. Department of Commerce, Bureau of the Census, 1991.

Statistical Abstract of the United States: 1992. Washington, DC: U.S. Department of Commerce, Bureau of the Census, 1992.

Stiglitz, Joseph E. "The Demand for Education in Public and Private School Systems." *Journal of Public Economics*, November 1974, *3*(4), pp. 349–85.

Summers, Anita A. and Wolfe, Barbara L. "Do Schools Make a Difference?" *American Economic Review*, September 1977, *67*(4), pp. 639–52.

***Wall Street Journal*.** [Blurb]. September 1993, p. 1.

______. "Private Groups Compete for the Chance to Create New Schools with Public Funds." January 24, 1994, p. B1.

Witte, John F. "School Choice and Student Performance," in Helen Ladd, ed., *Holding schools accountable*. Washington, DC: Brookings Institution, 1996, pp. 149–76.

Witte, John F.; Bailey, Andrea B. and Thomas, Christopher A. "Third-Year Report: Milwaukee Parental Choice Program." Mimeo, University of Wisconsin–Madison, December 1993.

Zimmerman, David J. "Regression Toward Mediocrity in Economic Stature." *American Economic Review*, June 1992, *82*(3), pp. 409–29.

The Social Selection of Flexible and Rigid Agents

By JOSEPH E. HARRINGTON, JR.*

People differ in how they respond to their environment. Some individuals treat each situation as unique and tailor their behavior accordingly while others respond in the same manner regardless of the situation. My objective is to explore how social systems select from such a heterogeneous population. A class of simple hierarchical systems is considered which encompasses some features of corporations and electoral systems. A selection process operates on this population which results in successful agents going on to compete against equally successful agents for further advancement. I characterize the population dynamics and the type of agent that ultimately dominates. (*JEL* D00, D23, D72)

Who will cure the nation's ills?
A leader with a selfless will.
But how will you find this leader of yours?
By a process of natural selection of course.

W. H. Auden (1940)

Across diverse circumstances, there is often regularity in the manner in which a person responds to his or her environment. This observed unity to a person's behavior is part of what defines that person. In the political arena, some politicians routinely support the popular positions on key issues while others, often referred to as ideologues, appear to embody a particular world view and support positions consistent with that view. The former are driven by the "here and now" of getting elected while the behavior of ideologues is rooted in those past events which determined their ideology and their commitment to that ideology.[1] Representative of such a contrast are Winston Churchill and Prime Minister Stanley Baldwin during the 1930's (William Manchester, 1988 p. 219):

> Winston was guided by a built-in gyroscope which would carry him toward his objective through tumult, while the prime minister relied on a kind of sociological radar—signals from voters—to determine his course.

A related form of heterogeneity is manifested in organizations. There are those individuals who strive to please their superior—commonly referred to as "yes men"—while others (let us call them mavericks) appear to possess a definite opinion as to what is the proper action, independent of their superior's opinions, and are committed to acting on it. "Yes men" are driven by the expedient

* Department of Economics, Johns Hopkins University, Baltimore, MD 21218. The paper has been improved by the comments and suggestions of many people, including Michael Alexeev, Larry Blume, Glenn Ellison, Tim Feddersen, Douglas Gale, Bruce Hamilton, Ken Kollman, Dean Lacy, Roger Lagunoff, Nolan McCarty, Scott Page, Tom Palfrey, Keith Scharfman, Ed Scheinerman, Joseph Schlesinger, Peter Streufert, three anonymous referees, seminar participants at the Wharton School, Johns Hopkins University, University of Maryland, Queens University, Indiana University, IUPUI, University of North Carolina, University of Toronto, University of Montreal, Ohio State University, University of Western Ontario, and University of Michigan (Program on Complex Systems), and conference participants at the Midwest Math Econ Meetings and the AEA Meetings. Let me also thank Bob Bradford and Chris Ruebeck for very able research assistance. Finally, I would especially like to thank Scott for taking the time to curl up under a palm tree and read the poetry of W. H. Auden. An earlier version of this paper was presented at the 1993 American Political Science Association Meetings and the 1994 Public Choice Society Meetings under the title "Electoral Selection and the Survival of Ideologues."

[1] See Aaron Wildavsky (1965) for a comparative discussion of ideologues and pragmatists and Gary M. Maranell (1970) for a ranking of U.S. presidents in terms of flexibility.

dictum "to get along, you go along," while mavericks "march to the tune of a different drummer." Turning to the ranks of businessmen, one finds that some are quick to adjust their business strategy to new market conditions while others dogmatically cling to their original plan.[2] A classic example is Henry Ford, whose stubborn commitment to the same strategy arguably caused Ford Motor Company to squander its leadership. Alfred P. Sloan, CEO for General Motors, commented (1965 pp. 163, 437):

> Mr. Ford, who had so many brilliant insights in earlier years, seemed never to understand how completely the market had changed ... [He] stayed too long with his old and once dominantly successful concept of the business.

An entrepreneur who was no stranger to change is Sam Walton (1992 p. 169):

> When folks buy into a way of doing things, and really believe it's the best way, they develop a tendency to think that's exactly the way things should always be done. So I've made it my own personal mission to ensure that constant change is a vital part of the Wal-Mart culture itself.

The commonalty to these diverse settings is that they all encompass people who differ in how they respond to a change in circumstances. Politicians who follow the dictates of their constituents, members of an organization who "go along," and entrepreneurs who routinely modify their business strategy are *flexible* in that their behavior is driven by the conditions of the environment—whether it is to win an election, get promoted, or increase profit. In contrast, political ideologues, organizational mavericks, and dogmatic entrepreneurs are *rigid* in that they appear to be endowed with some fundamental approach and rarely veer from it. Their behavior is relatively unresponsive to the current shape of the environment and is instead determined by past events that established the properties of their code of conduct.[3]

The central premise of this study is that a substantive source of heterogeneity among people is the degree to which they modify their behavior in response to a change in their environment. Indeed, there is a considerable psychology literature that documents and characterizes heterogeneity along a rigidity-flexibility dimension.[4] Let me now place such a diverse population within the confines of a government, corporation, or army. This leads to a central research question: given a population of flexible and rigid agents and a hierarchical social system within which they interact and compete, which type of agent will rise to the top? Are ideologues or office seekers more likely to occupy high-level offices like governor, senator, and president? Are apparatchiks or mavericks more proficient at climbing the corporate ladder and becoming senior vice presidents and chief executive officers? Are successful entrepreneurs more likely to have a business strategy built on persistence and dogma or flexibility and change? Are the leaders of our society more likely to be pragmatists who "blow with the wind" or visionaries with a "selfless will" who would sacrifice themselves as martyrs before compromising their beliefs?

Toward addressing these questions, I develop a simple class of hierarchical societies in which the population of agents differ in the flexibility of their behavior. My research objective is to understand how the properties of

[2] William D. Guth and Renato Tagiuri (1965) discuss how personal values might influence corporate strategy. At the level of business strategy, the trade-off between committing to an approach and committing to being flexible is the focus in Robert H. Hayes and Gary P. Pisano (1994).

[3] This type of heterogeneity manifests itself in the judicial realm as well, as was pointed out to me by Ken Scharfman. A rigid judge may be interpreted as one who acts in "good faith" in that he administers his position based "on the reasons provided by the law" even though he may personally disagree with the decision or know it makes society worse off (Steven J. Burton, 1992 pp. 36–37). In opposition is the expedient judge who fails to exercise self-restraint and lets personal interests and preferences, rather than overriding principles, influence decisions.

[4] See, for example, Abraham S. Luchins and Edith Hirsch Luchins (1959), William A. Scott (1966), David H. Jonassen and Barbara L. Grabowski (1993).

a social system influence the characteristics of those agents who do well within that system. This research fits into a growing line of work which uses an evolutionary-style framework to explore the types of preferences or behavioral rules that survive in a competitive environment. This work is unique in modelling the social environment as hierarchical and in having agents differ in terms of their flexibility. A similar form of heterogeneity is considered in John Haltiwanger and Michael Waldman (1991), Robert W. Rosenthal (1993), and Dale O. Stahl (1993) and, at a population level, in Elchanan Ben-Porath et al. (1993). Encompassing distinct forces, Ronald A. Heiner (1983) investigates the appropriate amount of flexibility. The flexibility of behavioral rules is also central to the time-consistency problem in macroeconomics and is the root of the debate on rules versus discretion; see Finn E. Kydland and Edward C. Prescott (1977). From this perspective, this paper questions whether agents endowed with a rule survive at a higher or lower rate than those endowed with discretion.

I. A Class of Hierarchical Competitive Systems

I start with a hierarchical system with a lowest level and no upper bound on the highest level. At each level there is a large population of agents, specifically, a continuum of agents.[5] Whether it is a politician trying to advance from the state legislature to the House of Representatives or a regional manager striving to become a vice president in a corporation, advancement typically requires performing relatively better than some subset of peers who are also eligible for promotion. This process is modelled by assuming that, at each level, agents are randomly matched into pairs and compete for "promotion" to the next level.[6] The terms promotion, advancement, and survival interchangeably are used. The assumption that an agent is compared to only one other agent is a concession to tractability but would seem to be reasonable for electoral systems where general elections typically involve two candidates. Each of these matchings is faced with a stochastic environment. Once the environment is determined and revealed to the agents, they choose actions. The agent with greater performance is promoted to the next level while the other agent is assumed to drop out of the system or, more to the point, no longer be eligible for promotion. While this "up-or-out" structure is extreme, it is not without merit. Casual observation suggests that a large percentage of candidates who lose do not run again and corporate employees who are "passed over" when their time has come may no longer be on the "fast track," which makes them less likely to be considered for promotion.[7]

There are two possible environments which I denote type 0 and type 1. At each level, a proportion b of all matchings have a type 1 environment. This is assumed to be i.i.d. across levels so that the probability an agent faces a type 1 environment is b and this is independent of his personal history. While each agent faces an uncertain future environment, the absence of aggregate uncertainty simplifies matters. Without loss of generality, I make a type 1 environment more common: $b \in (1/2, 1)$.[8]

In responding to one's environment, there are two generic approaches. Depending on the context, they could correspond to a political ideology, a corporate ethic, a business strategy, or yet some other concept. In this simple setup, an approach or a strategy corresponds

[5] Assuming a continuum of agents simplifies the analysis by making the process deterministic. Richard T. Boylan (1992) shows that a deterministic system can be a valid approximation for a stochastic system when there are many agents.

[6] It would be interesting to allow for correlated matching as it might simulate self-selection among agent types. This would seem applicable to electoral systems where the decision to run for higher office depends on one's prospective opponent.

[7] Joseph A. Schlesinger (1966, 1991) documents the progressive paths taken to higher office. The tournament-style structure of organizations is examined in Sherwin Rosen (1986), while Raaj K. Sah and Joseph E. Stiglitz (1991) also explore the determinants of upper-level management.

[8] The case of $b = 1/2$ turns out to be knife-edge. When $b \neq 1/2$, there is a finite number of rest points. When $b = 1/2$, there is a continuum of rest points and at most one of them is locally stable. Details are in Harrington (1994).

to a particular action to play in all environments. By definition, action 0 (1) is the action associated with strategy 0 (1).[9] In a manner to be described momentarily, action 0 (1) is the best action for a type 0 (1) environment. Since $b > 1/2$, action 1 is more frequently the appropriate response to the environment.

Selection is determined as follows. If the two matched agents choose distinct actions, then the agent whose action matches the environment survives and is promoted to the next level. If both agents select the action which matches the environment, the agent who has chosen that particular action more frequently in the preceding h rounds advances with probability $p \in [1/2, 1]$. If they have chosen that action equally frequently, then an agent is randomly selected to survive. There is no need to specify whom is promoted if both agents choose the less appropriate action as the set of equations which describe the population dynamics is independent of it.

The idea is that survival depends on current performance which itself is determined by one's action and one's proficiency with that action where proficiency comes from experience. If $p > 1/2$, then experience yields an advantage which could be due to learning-by-doing or, as in the electoral context, credibility that comes from being relatively consistent in one's positions over time. If $p = 1/2$, then there is no experiential advantage. p is a measure of how much experience matters.[10] Note that survival depends lexicographically on one's current action and one's experience with that action. This means that the incremental effect from choosing a better action exceeds the incremental effect from more experience. h is a parameter which determines how much of an agent's history is relevant for proficiency. I initially consider the case of $h = \infty$ so that an agent is more effective in using a particular action if he has chosen that action more often over the entire history of play. I then examine the case when $h = 1$ so that only the most recent past matters. The value of h could be determined by the rate of depreciation of knowledge regarding the proper use of an action or, in the electoral context, by memories of voters. Having advancement depend only on current performance is like the old Hollywood adage, "You're only as good as your last picture," or voters' lament of "What have you done for me lately?" and obviously implies a certain myopia or forgetfulness. This might be a reasonable assumption for the electoral context but is admittedly uncompelling for the corporate setting. In Section III, past performance is allowed to play a limited role.

Each agent is endowed with a behavioral rule. The space of agent types is then associated with the space of feasible behavioral rules. Until Section III, attention is limited to behavioral rules that condition only on the current environment. It is then the set of functions which map the set of environments, {0, 1}, into the set of actions, {0, 1}, with the exception that the pathological case of always choosing an action inappropriate for the current environment is excluded. I believe this simplifies the analysis without any loss of generality. A flexible agent is defined to be one who always selects the action best suited for the environment: he chooses action 0 (1) when the environment is type 0 (1). A rigid agent chooses the same action irrespective of the environment. A type 0 rigid agent always uses action 0 and a type 1 rigid agent always uses action 1.

In concluding, there are a variety of sources of heterogeneity in rigidity. Faced with a complex meta-environment and computational constraints, people may simply differ in their opinion as to which behavioral rule is best. Alternatively, some agents may be endowed with a greater ability to modify their behavior. Perhaps more flexible agents can more finely discriminate among different environments. An agent who cannot tell the difference between two environments cannot condition his behavior on which environment occurs.[11] Th

[9] Since there is a one-to-one mapping between strategy sets and action sets, there is no formal distinction between the two. Conceptually, there is a distinction in that a strategy is a rule which maps from the space of environments into the space of actions. It just so happens these strategies call for the same action for all environments.

[10] The results also extend to $p < 1/2$, but that range is inconsistent with the notion of experience contributing to proficiency.

[11] For the relation between an organism's representational system and his action set, see Derek Bickerton (1990).

following argument for constrained choice sets is also relevant (Heiner, 1983 p. 585):

> ... agents cannot decipher all of the complexity of the decision problems they face, which literally prevents them from selecting most preferred alternatives. Consequently, the flexibility of behavior to react to information is constrained to smaller behavioral repertoires that can be reliably administered.

It is then quite plausible that people with different powers of reasoning, recognition, and memory would have different behavioral repertoires.

II. The Unbounded Memory System

The initial population resides at the lowest level of the system and is comprised of three types: rigid agents endowed with action 1, rigid agents endowed with action 0, and flexible agents. The objective is to understand how the characteristics of this cohort of agents evolves as it migrates up through the hierarchy.[12] Though the system begins with only three distinct types, there are potentially many types after the population has a chance to interact with the environment because an agent's performance depends not only on his endowed behavioral rule but also his personal history. For present purposes, it is sufficient to partition the population as follows.

Definition: Let $z_s^t \in \{0, 1\}$ denote the action of agent s at level t. For an h-memory system, agent s is **proficient in action** i at level t if $z_s^\tau = i$ for all $\tau \in \{\max\{1, t-h\}, \ldots, t-1\}$.

$r_i^t \equiv$ proportion of the level t population that are rigid agents endowed with action i (Ri); $i \in \{0, 1\}$.

$f_i^t \equiv$ proportion of the level t population that are flexible agents who are proficient in action i (Fi); $i \in \{0, 1\}$.

$x^t \equiv$ proportion of the level t population that are flexible agents who are not proficient in any action (FN).

The level t state of the system is defined by $(r_1^t, f_1^t, r_0^t, f_0^t, x^t)$. It is assumed that $r_1^1 > 0$, $r_0^1 \geq 0$, and $1 - r_1^1 - r_0^1 > 0$.

In this section, I consider the case of unbounded memory ($h = \infty$) so that an agent's entire history matters. For $t \geq 2$, the state at level $t+1$ is described by the following system of equations:[13]

(1) $r_1^{t+1} = 2r_1^t[(1/2)r_1^t + (b/2)f_1^t + br_0^t + bp(1 - r_1^t - f_1^t - r_0^t)]$,

(2) $f_1^{t+1} = 2f_1^t[(b/2)r_1^t + (b/2)f_1^t + br_0^t + bp(1 - r_1^t - f_1^t - r_0^t)]$,

(3) $r_0^{t+1} = 2r_0^t[(1/2)r_0^t + ((1 - b)/2)f_0^t + (1 - b)r_1^t + (1 - b)p(1 - r_1^t - r_0^t - f_0^t)]$,

(4) $f_0^{t+1} = 2f_0^t[((1 - b)/2)r_0^t + ((1 - b)/2)f_0^t + (1 - b)r_1^t + (1 - b)p(1 - r_1^t - r_0^t - f_0^t)]$,

where I have substituted $1 - r_1^t - f_1^t - r_0^t - f_0^t$ for x^t. In each of these equations, the bracketed term is the probability that an agent of that type survives. Multiplying this by the level t proportion of that type and doubling it (because an agent could be the first or second draw in a matching) gives a type's proportion at the next level. Consider randomly matching two agents and suppose the first agent drawn is an $R1$. With probability r_1^t he is matched with the same type, in which case his probability of surviving is $1/2$. With probability f_1^t he is matched with an $F1$, in which case he survives with probability $1/2$ when the environment is type 1, which occurs with probability b, and does not survive when the environment is type 0 as an $F1$ selects action 0. With probability r_0^t, he is matched with an $R0$ and survives when the environment is type 1. Finally, if he is matched with either an $F0$ or an FN,

[12] If the hierarchy is to be kept "full" then, at the end of each round, a fresh cohort of agents must enter the lowest level to replace those who moved on. This is not relevant to the current analysis in that I am interested in what happens to a single cohort. In Harrington (1996), a flow of new agents is modelled where a new agent's behavioral rule is assumed to be partly endowed and partly the result of imitation of those at higher levels.

[13] The system of equations which determines $(r_1^2, f_1^2, r_0^2, f_0^2, x^2)$ is different from that in (1)–(4) because agents do not have a history at level 1.

which occurs with probability $(1 - r_1^t - f_1^t - r_0^t)$, and the environment is type 1, he survives with probability p because of his greater experience with action 1. In a similar fashion, the other three equations can be explained.

Theorem 1 establishes that a population with all $R1$s is globally stable when the experiential advantage is sufficiently great and/or the type 1 environment is sufficiently common. All proofs are in the Appendix.

THEOREM 1: *With unbounded memory* ($h = \infty$), *if* $pb > 1/2$ *then* $\lim_{t \to \infty} r_1^t = 1$.

Regardless of their initial presence in the population, systems with sufficiently many levels find their highest levels dominated by rigid agents who use the action that is more frequently the appropriate response to the environment. Rigid agents endowed with the less effective action and flexible agents are eventually eliminated.[14]

Theorem 1 is best understood by examining the system's dynamics. It is straightforward to show that the presence of agents who are proficient in action 0 is always diminishing: $\Delta r_0^t < 0$ and $\Delta f_0^t < 0$, where $\Delta r_i^t \equiv r_i^{t+1} - r_i^t$ and $\Delta f_i^t \equiv f_i^{t+1} - f_i^t$. Turning to those agents who are proficient in action 1, one can use (1)–(2) to derive:

$$(5) \quad \text{If } r_1^t > 0 \text{ then: } \Delta r_1^t \gtreqless 0 \text{ as } f_1^t \lesseqgtr [(2pb - 1)(1 - r_1^t) + 2b(1 - p)r_0^t]/b(2p - 1).$$

$$(6) \quad \text{If } f_1^t > 0 \text{ then: } \Delta f_1^t \gtreqless 0 \text{ as } r_1^t \lesseqgtr [((2pb - 1) + 2b(1 - p)r_0^t)/b(2p - 1)] - f_1^t.$$

Though $F1$s are eventually driven to extinction, their presence is increasing when the proportion of $R1$s is sufficiently small. Though $R1$s are eventually driven to domination, they are decreasing when the proportion of $F1$s is sufficiently great. Note that when $r_0^t = 0$ (or $p = 1$), (5)–(6) depend only on (r_1^t, f_1^t) and Figure 1 depicts the associated phase diagram. The use of Figure 1 to explore the dynamics is reasonable since $R0$s are monotonically decreasing and simulations reveal that they rapidly converge to zero.

Suppose the system currently has many $R1$s and $F1$s so that it is initially in region I. For example, when $p \cong 1$ and $r_1^1 \cong r_0^1$, it can be shown that the level 2 state of the system is in region I. Let me describe the resulting dynamical path, an example of which is depicted in Figure 2. The large presence of $F1$s means that $R1$s are frequently meeting agents who are equally experienced in action 1 but are superior by means of their flexibility. As a result, the proportion of $R1$s is shrinking since they survive at a rate of only $b/2$ ($<1/2$) in their encounters with $F1$s. Due to the large presence of $R1$s in the population, the relative advantage of an $F1$ is not his proficiency in action 1 (since that is matched by $R1$s) but rather his flexibility. Hence, $F1$s tend to do best when the environment is type 0 as then, by choosing action 0, they capitalize on $R1$s being locked into action 1. Though $F1$s are surviving at a high rate, they are losing proficiency in action 1. This results in the proportion of $F1$s shrinking and the proportion of FNs growing (see Figure 2). Since the presence of $R1$s is diminishing in region I and the presence of $R0$s is always diminishing, the presence of flexible agents is growing in region I. When the proportion of $F1$s becomes sufficiently small, the system moves into region II. At that point the population is favorable to $R1$s in that they are frequently meeting agents who are not proficient in action 1, in which case their survival rate is at least pb ($>1/2$). From then on the proportion of $R1$s grows. For this path, the population dynamics are nonmonotonic in that the presence of flexible agents is higher at intermediate levels than at low and high levels.

Alternatively, consider when the system begins in region III. Due to their small presence, $R1$s and $F1$s are largely meeting agents who are not proficient in action 1. Hence, $R1$s and $F1$s are thriving. Note, in particular, that the

[14] It is shown in Harrington (1994) that (r_1, f_1, r_0, f_0, x) is a rest point if and only if $(r_1, f_1, r_0, f_0, x) \in \{(1, 0, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 0, 1), (0, (2b - 1)/b, 0, 0, (1 - b)/b)\}$.

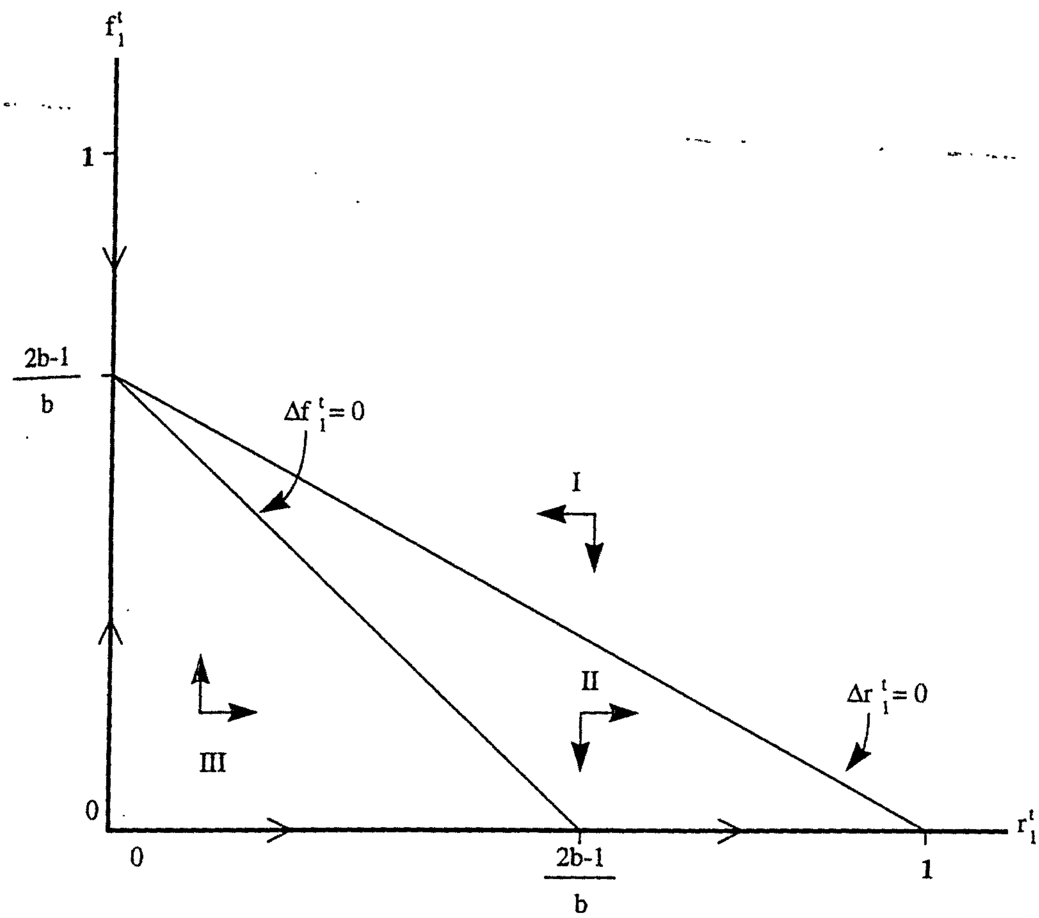


FIGURE 1. UNBOUNDED MEMORY SYSTEM: PHASE DIAGRAM (ASSUMPTION: $r_0^t = 0$ or $p = 1$)

proportion of $F1$s is growing because their differential advantage against an $R0$, $F0$, or FN is being proficient in action 1 rather than being flexible. Thus, their survival rate is highest when the environment is type 1, in which case they choose action 1, survive with probability at least p, and maintain their proficiency in action 1. With $F1$s and $R1$s growing, eventually they start frequently meeting each other. Given their frequent encounters with $R1$s, the differential advantage of an $F1$ shifts to being flexible, in which case their presence shrinks as they convert to FNs. This is represented by the system moving into region II. At that point, the presence of $R1$s is sufficiently great that they continue to grow and that growth accelerates as the presence of $F1$s shrinks. An interesting property of this path is that the proportion of $F1$s is growing at low levels even though they eventually become extinct.

While $R1$s can initially founder, they start thriving once the proportion of $F1$s is sufficiently small. What is surprising is how long this can take. Since a flexible agent must have faced $T - 1$ consecutive type 1 environments to be proficient in action 1 at level T, one would expect there to be very few proficient flexible agents when T is not small and b is not close to one since the likelihood of $T - 1$ consecutive type 1 environments is b^{T-1}. It would then seem that $R1$s would quickly overwhelm $F1$s. What this ignores, however, is that those flexible agents who are lucky enough to be proficient in action 1 have a much higher chance of surviving by virtue of their proficiency. For example, suppose $(b, p, r_1^1, r_0^1) = (0.6, 1, 0.25, 0.25)$, as in Figure 2. Of the proficient agents, those who are flexible are on the order of 1 in every 2 agents at level 10, 1 in every 3 at level 15, 1 in every 4

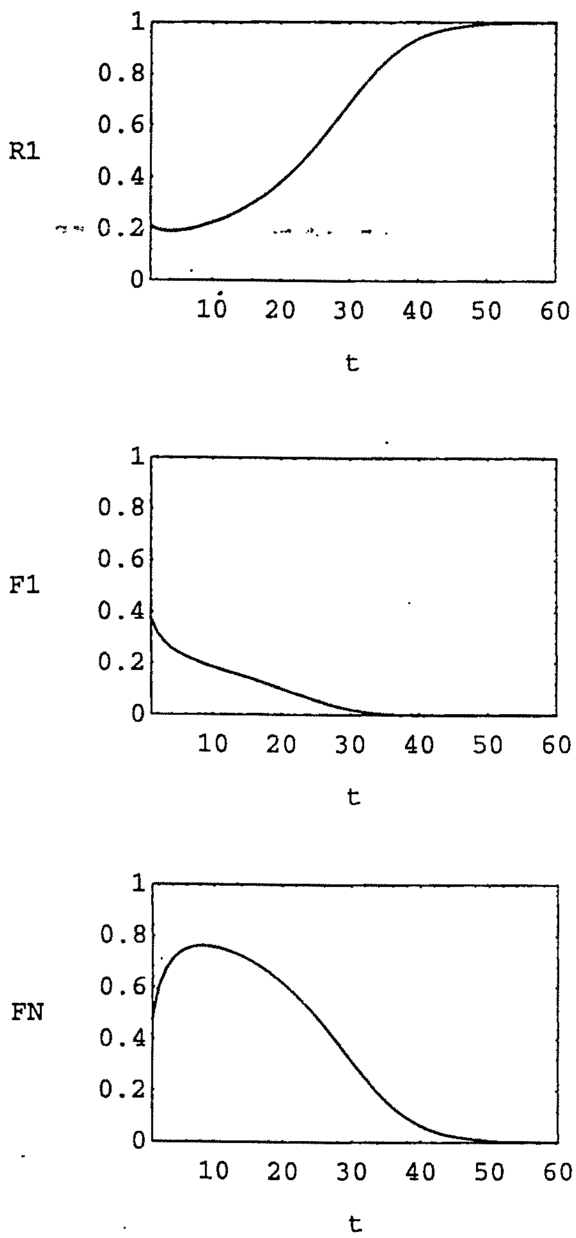


FIGURE 2. UNBOUNDED MEMORY SYSTEM: EXAMPLE ($b = 0.6, p = 1, r_1^1 = 0.25, r_0^1 = 0.25$)

at level 20, and 1 in every 8 at level 25. By comparison, when selection is inoperative so that all agents survive, of all proficient agents those who are flexible are around 1 in every 100 at level 10, 1 in every 1,000 at level 15, 1 in every 15,000 at level 20, and 1 in every 200,000 at level 25. Thus, selection allows proficient flexible agents to survive for many more rounds. By maintaining proficiency, flexible agents may be able to limit the dominance of rigid agents even in systems with relatively many levels.

Now suppose experience with an action is sufficiently unproductive in improving survival. Theorem 2 establishes that flexible agents will dominate.

THEOREM 2: *With unbounded memory* ($h = \infty$), *if* $pb \leq 1/2$ *then* $\lim_{t \to \infty} (r_1^t + r_0^t) = 0$.

Since, when the system is run long enough, most flexible agents are not proficient in any action, the long-run dominance of $R1$s is contingent upon them doing well against such agents. Given that when an $R1$ meets such an agent his probability of survival is pb, their long-run survival requires $pb > 1/2$. Otherwise, as shown in Theorem 2, flexible agents will dominate.

III. Extensions of the Unbounded Memory System

The objective of this research is not to determine whether, in general, rigid behavior is prevalent but rather to understand how the properties of a social system relate to the type of behavior that thrives within that system. Contrary to a common view that flexibility is a superior trait, the analysis of the previous section identified one class of social systems in which rigid agents do quite well. I now want to consider related systems so as to both assess the robustness of this result and to understand what properties promote rigid behavior. To simplify proofs, I assume $p = 1$ for the remainder of the analysis.

A. *More Sophisticated Flexible Agents*

Whenever one engages in an analysis that assumes agents are endowed with behavioral rules, it is important to ask whether there is a behavioral rule not in the feasible set that would have thrived. Along these lines, I go beyond strategies that condition only on the current environment and consider making flexible agents a bit more sophisticated. The preceding analysis suggests that a flexible agent, when faced with a rigid agent, might want to mimic a rigid agent. This is embodied in the following behavioral rule.

(7) At level 1, choose the action that matches the environment. At level t (≥ 2), choose the action that matches the environment except when:

(i) one has always chosen action i;
(ii) the current environment is type j ($\neq i$); and
(iii) one is matched with an agent who is type Ri; $i \in \{0, 1\}$.

Compared to the rule of always choosing the best action for the current environment, this rule enhances proficiency while maintaining a chance of surviving today though the probability of that event is lower.

THEOREM 3: *With unbounded memory* ($h = \infty$), *if flexible agents use the rule in* (7) *then* $\lim_{t\to\infty} r_1^t = 1$.

While this more sophisticated rule may delay the time at which an flexible agent loses proficiency, such remains inevitable. Interestingly, since $F1$s no longer take advantage of their flexibility when facing $R1$s, the proportion of $R1$s is now monotonically increasing for all initial conditions (see the proof of Theorem 3 in the Appendix).[15]

A related possibility to (7) is for a proficient flexible agent to act "rigid" when facing any identically proficient agent. If all flexible agents use such a rule then one maintains a chance of surviving today and lengthens the time for which one is maximally proficient.[16]

(8) At level 1, choose the action that matches the environment. At level t (≥ 2), choose the action that matches the environment except when:

(i) one has always chosen action i;
(ii) the current environment is type j ($\neq i$); and
(iii) one is matched with an agent who has always chosen action i; $i \in \{0, 1\}$.

THEOREM 4: *With unbounded memory* ($h = \infty$), *if flexible agents use the rule in* (8) *then* $\lim_{t\to\infty} (r_1^t + f_1^t) = 1$.

While flexible agents can now survive in the long run, this is only achieved by perfectly mimicking rigid agents! Since, in the limit, $F1$s face only other $F1$s and $R1$s, it follows from (8) that they choose action 1 regardless of the environment. Furthermore, if they ever consider deviating from this rule and choosing action 0 when the environment is type 0, it is known by the argument in the preceding section that they will eventually be driven out.

B. *Survival Depends on Past Performance*

Thus far, survival depends only on current performance which itself depends on an agent's action and his experience with that action. In many settings, past performance would seem quite relevant. For example, promotion in a corporation is apt to depend on one's performance throughout one's tenure, though perhaps with disproportionate weight placed on recent results.[17] To encompass past performance, consider the following selection rule.

(9) The agent who chose the best action for the current environment survives.
If both agents chose the best action then the one who more frequently in the past chose the best action for the environment at that time survives.
If both agents chose the best action and equally frequently in the past chose the best action for the environment at that

[15] An agent who only knows the history of environments and actions of the other agents may not be able to implement the rule in (7). For example, if another agent has always chosen action 1 and always faced a type 1 environment then he could be either rigid or flexible. The rule in (7) is then a best-case scenario for flexible agents in that it presumes more information than they realistically would have.

[16] This rule can be implemented with information only on another agent's past actions.

[17] A corporation might choose to focus on current performance under the presumption (perhaps false) that one's past performance must have been reasonably good for the agent to have advanced to the current level.

time, then the one who chose the current action more frequently in the past survives.

Otherwise, let me retain the specification in Section II. Survival depends lexicographically on the current action, past performance, and experience. Given that flexible agents always choose the best action and always have maximal past performance, encompassing past performance results in flexible agents thriving.

THEOREM 5: *With unbounded memory* ($h = \infty$) *and the selection rule in* (9), $\lim_{t\to\infty}(r_1^t + r_0^t) \in ((2b-1)r_1^1/b(1-r_0^1),\ r_1^1 + r_0^1)$.

Since $\lim_{t\to\infty}(r_1^t + r_0^t) < r_1^1 + r_0^1$, then this social system favors flexible agents. However, rigid agents are not driven out. For example, when $b = 0.75$ and the initial population is equally divided between $R1$s, $R0$s, and flexible agents, rigid agents make up around 36 percent of the highest levels in systems with sufficiently many levels.

To explain how rigid agents survive, let $R1^*$ denote a rigid agent endowed with action 1 who has always faced a type 1 environment. The notable property of an $R1^*$ is that he has always chosen the action which is best for the environment, just like a flexible agent. By the selection criterion in (9), survival of rigid agents rests upon $R1^*$s doing well. This may not seem very promising because it would appear unlikely for any agent to continually face the same environment. Indeed, for the selection criterion in Section II, one can show that the proportion of agents who are $R1^*$s goes to zero as $t \to \infty$. However, that selection criterion provided no pressure for $R1$s to accumulate such a history. In contrast, given the selection criterion in (9), $R1$s who just happen to face a type 1 environment in, say, the first two rounds, have a higher chance of surviving then $R1$s who have not. In particular, when faced with any type other than an $F1$, the survival rate of an $R1^*$ (as an $R1^*$) is b (which exceeds 1/2) in that he survives for sure when the environment is type 1. For example, against a flexible agent who has not always chosen the same action, an $R1^*$ survives when the environment is type 1, as he has the advantage of having always chosen the same action [the third part in (9)] and, like any flexible agent, has always chosen the action right for the environment including the current one. Against an $F1$, however, an $R1^*$'s survival rate is only $b/2$ ($<1/2$) since, when the environment is type 1, the two agent types have identical histories and choose identical actions. Given that long-run survival requires a survival rate of at least 1/2, the long-run survival of $R1^*$s (and thus rigid agents) is then contingent upon there not being too many $F1$s. What assures that there are not too many $F1$s is that the dynamic affecting $F1$s is exactly the same as that for $R1^*$ since the two agent types are identical (and, as a result, the ratio of $R1^*$s to $F1$s is constant; see the proof of Theorem 5 in the Appendix). Therefore, $F1$s and $R1^*$s shrink together and eventually stabilize at levels above zero. In particular, one can show that the proportion of agents who have always faced a type 1 environment—$R1^*$s and $F1$s—converges to $(2b-1)/b$ as $t \to \infty$ with the split between them depending on the level in the population.[18] In short, while it is relatively unlikely for any specific agent to exclusively face a type 1 environment, some agents will have done so and selection results in them being disproportionately represented in the next generation. In this manner, the unlikely—a rigid agent having always chosen the action best for the environment—is made likely by selection pressures and results in rigid agents surviving.

I find that allowing selection to depend on past performance is conducive to flexible agents dominating though rigid agents still manage to survive. Finally, let me note that if the third part of (9) were replaced with a random selection of agents, there would be no virtue to being rigid. In that case, rigid agents would be driven out and only flexible agents would be present.[19]

[18] Since, the proportion of $R1^*$s is positive as $t \to \infty$, this follows immediately from (A10).

[19] If the third part in (9) was replaced with the flip of a fair coin then $\Delta\hat{r}_1^t = \hat{r}_1^t[(2b-1) - (\hat{r}_1^t + f_1^t)]$, where $\hat{r}_1^t$ denotes the proportion of $R1^*$ agents. Since $\hat{r}_1^t >$ ... implies $\Delta\hat{r}_1^t < 0$, it follows that $\lim_{t\to\infty}\hat{r}_1^t = 0$.

IV. The Bounded Memory System

An implication of assuming the system has unbounded memory is that a flexible agent who becomes less experienced than a rigid agent is always less experienced. Since such an assumption is clearly extreme and could be crucial for the domination of rigid agents, I now consider a bounded memory system while otherwise maintaining the structure of Section II (except for assuming $p = 1$). As I am particularly interested in assessing the robustness of my earlier results, the extreme case of a one-period memory is assumed. Proficiency is then determined solely by an agent's behavior in the preceding round and, in principle, is relatively easy to achieve.

With a one-period memory, all agents are proficient so that there are four types: $\{R1, F1, R0, F0\}$. The dynamical system for $t \geq 2$ is:

$$r_1^{t+1} = 2r_1^t[(1/2)r_1^t + (b/2)f_1^t + br_0^t + bf_0^t], \tag{10}$$

$$f_1^{t+1} = 2f_1^t[(b/2)r_1^t + (b/2)f_1^t + br_0^t + bf_0^t] + 2f_0^t[br_0^t + (b/2)f_0^t], \tag{11}$$

$$r_0^{t+1} = 2r_0^t[(1-b)r_1^t + (1-b)f_1^t + (1/2)r_0^t + ((1-b)/2)f_0^t], \tag{12}$$

$$f_0^{t+1} = 2f_0^t[(1-b)r_1^t + (1-b)f_1^t + ((1-b)/2)r_0^t + ((1-b)/2)f_0^t] + 2f_1^t[(1-b)r_1^t + ((1-b)/2)f_1^t]. \tag{13}$$

In (11), $2f_0^t[(b/2)f_0^t + br_0^t]$ represents those agents who switch from being an $F0$ to being an $F1$, while in (13) $2f_1^t[(1-b)r_1^t + ((1-b)/2)f_1^t]$ captures the flow of $F0$s to $F1$s. In contrast, with unbounded memory, the only flow was from $F1$s and $F0$s to FNs.

Though the bounded memory system does not have a global attractor, it does have asymptotic attractors. Let $d((r_1', f_1', r_0', f_0'), (r_1'', f_1'', r_0'', f_0''))$ be the Euclidean distance between states (r_1', f_1', r_0', f_0') and $(r_1'', f_1'', r_0'', f_0'')$.

Definition: (r_1, f_1, r_0, f_0) is an **asymptotic attractor** if there exists $\varepsilon > 0$ such that if $d((r_1^1, f_1^1, r_0^1, f_0^1), (r_1, f_1, r_0, f_0)) < \varepsilon$ then $\lim_{t\to\infty}(r_1^t, f_1^t, r_0^t, f_0^t) = (r_1, f_1, r_0, f_0)$.

A state is an asymptotic attractor if the system converges to it when it is in a sufficiently small neighborhood of it. This is then a kind of local stability.

It is shown below that the bounded memory system has two asymptotic attractors. Given the ease with which proficiency is achieved, one has all flexible agents. More interestingly is that the other has all rigid agents (endowed with action 1).[20]

THEOREM 6: *With bounded memory ($h = 1$), (r_1, f_1, r_0, f_0) is an asymptotic attractor if and only if $(r_1, f_1, r_0, f_0) \in \{(1, 0, 0, 0), (0, b, 0, 1-b)\}$.*

To explore the system's dynamics, I construct a phase diagram under the assumption that $r_0^t = 0$. As with unbounded memory, the proportion of $R0$s monotonically converges to

[20] A qualification to Theorem 6 is in order. By definition, a rest point is not an asymptotic attractor if within any neighborhood around the rest point there are states such that the system does not converge to it. This leaves open the possibility that there could be other states within the neighborhood for which the system does converge. While this is not true for the rest point with all $R0$s, I have been unable to prove that there do not exist convergent paths for the rest point with a mix of flexible and rigid agents though, after an extensive numerical search, no convergent paths were found. Note, however, that if the limit is taken so as to make this a continuous time system, this rest point would be a saddlepoint which would imply that almost all paths would lead away from it.

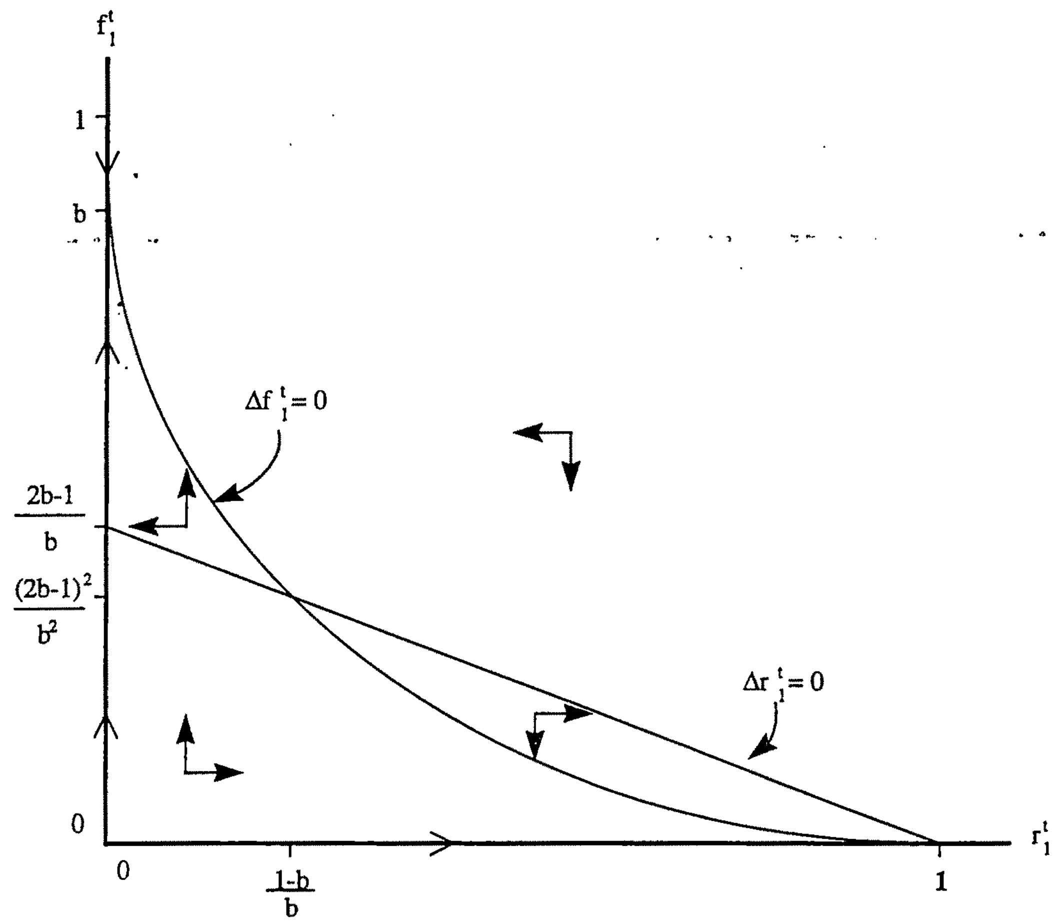


FIGURE 3. BOUNDED MEMORY SYSTEM: PHASE DIAGRAM (ASSUMPTION: $r_0^t = 0$)

zero and simulations show that the rate of convergence is rapid. Given $r_0^t = 0$ and substituting $1 - r_1^t - f_1^t$ for f_0^t in (10)–(11), we derive:

$$(14)\quad \text{If } r_1^t > 0 \text{ then: } \Delta r_1^t \gtreqless 0 \text{ as } f_1^t \lesseqgtr [(2b-1)/b](1-r_1^t) \equiv \Gamma(r_1^t),$$

$$(15)\quad \text{If } f_1^t > 0 \text{ then: } \Delta f_1^t \gtreqless 0 \text{ as } f_1^t \lesseqgtr [b(1-r_1^t)^2/(1-br_1^t)] \equiv \Omega(r_1^t).$$

$\Gamma(\cdot)$ is the same as when the system has unbounded memory (assuming $p = 1$) while $\Omega(\cdot)$ is distinct.

$$(16)\quad \Omega(0) = b > (2b-1)/b = \Gamma(0);$$

$$\Omega(1) = 1 = \Gamma(1).$$

$$\Omega'(r_1^t) = -[b(1-r_1^t)/(1-br_1^t)^2] \times [(1-b) + (1-b(r_1^t)^2)] < 0 \ \forall r_1^t \in [0,1);\ \Omega'(1) = 0.$$

$$\Gamma'(r_1^t) = -[(2b-1)/b] < 0 \quad \forall r_1^t \in [0,1].$$

Using the properties in (16), the phase diagram is constructed in Figure 3. The two asymptotic attractors are represented by $(r_1, f_1) = (1, 0)$ and $(r_1, f_1) = (b, 0)$. $(r_1, f_1) = ((1-b)/b, (2b-1)^2/b^2)$ represents a rest point that is not asymptotically stable.

First note that when the proportion of $R1$s is sufficiently great, the population dynamics are qualitatively similar to when memory is unbounded (see Figure 1). This is interesting in light of the qualitative distinction between unbounded and bounded memory systems. When memory is unbounded, proficiency in the better action requires a flexible agent to have always faced a type 1 environment (and played action 1). Flexible agents are then almost surely less experienced than rigid agents when the system has sufficiently many levels. In the bounded memory system, a flexible agent who chooses action 1 becomes proficient in action 1 *regardless* of his preceding history. In spite of this difference, bounding the memory does not qualitatively affect the population dynamics as long as the initial presence of $R1$s is sufficiently great.

To explain this finding, recall that flexible agents can continuously gain and lose proficiency in action 1 when memory is bounded. An $F1$ faced with a type 0 environment at level t will become an $F0$ (due to having chosen action 0) but switches back to being an $F1$ at level $t + 1$ if the environment at that time is type 1. The difficulty, however, lies in that an agent is unlikely to survive long enough to regain proficiency in action 1. With a strong presence of $R1$s, a flexible agent is very likely to meet an $R1$ who will take advantage of this flexible agent's "temporary" lack of proficiency in action 1. Rigid agents can then dominate even when it only takes one period for a flexible agent to achieve comparable proficiency to that of a rigid agent.

When instead the initial proportion of rigid agents is small, Figure 3 shows that the population dynamics are very different from when the system's memory is unbounded. Even though $R1$s can grow (when there are sufficiently few $F1$s), eventually flexible agents dominate as the system converges to $(r_1, f_1, r_0, f_0) = (0, b, 0, 1 - b)$. To understand the long-run domination of flexible agents, one must understand how $R1$s are driven out. As described above, the key to $R1$s surviving is that flexible agents who lose their proficiency in action 1 have a difficult time regaining it. However, when there are sufficiently few $R1$s in the population, a flexible agent who loses proficiency in action 1 has a reasonable chance

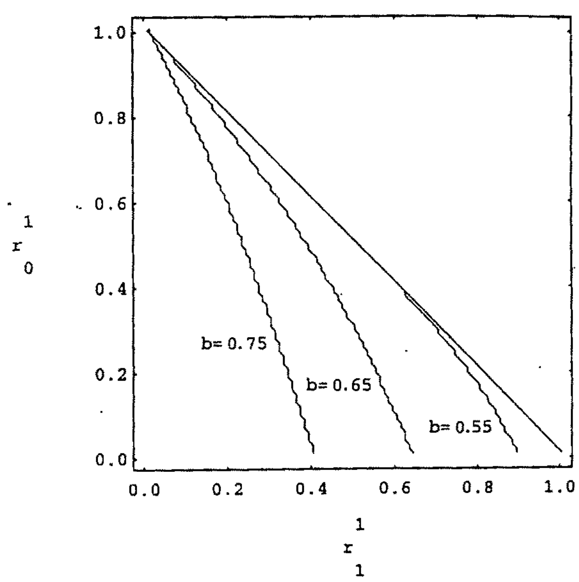


FIGURE 4. BOUNDED MEMORY SYSTEM: BASINS OF ATTRACTION

of avoiding an $R1$. This gives a flexible agent the opportunity to successfully regain proficiency in action 1. In particular, this is achieved by an $F0$ meeting his own kind. With this steady replenishment, the proportion of $F1$s is consistently high when there are initially few $R1$s. By facing a perpetually inhospitable population, $R1$s are driven out so that flexible agents flourish.

To characterize the basins of attraction of the two asymptotic attractors, numerical methods were used. In contrast to Figure 3, I allow $r_0^1 \neq 0$. The basins are defined over the space of initial populations where an initial population is represented by (r_1^1, r_0^1). Simulations showed that the basins partitioned the initial state space. These results are shown in Figure 4 where all points between a jagged line (where the associated value for b is noted) and the diagonal comprise the basin for the asymptotic attractor with all rigid agents, and all points to the left of a jagged line comprise the basin for the asymptotic attractor with all flexible agents. Note that if the initial population is biased towards flexible agents then this bias is magnified so that eventually nearly every surviving agent is flexible. When instead the initial population is biased towards rigid agents then the system converges to the asymptotic attractor with all rigid agents. As one

increases the frequency with which action 1 is the appropriate response to the environment, the basin for the rest point with all rigid agents expands. If action 1 is made more effective then it is more likely that agents who exclusively use that action will dominate. This result can also be interpreted as follows. As b gets closer to $1/2$, the environment becomes more volatile in which case being flexible is relatively more advantageous. Hence, flexible agents are more likely to dominate in that case.

V. Concluding Remarks

Environments routinely change and with them changes what is required to be well adapted. The extent of behavioral response to an altered set of surroundings is a crucial form of variation among organisms in the animal kingdom and would seem central to the determination of which organisms thrive and which perish. In spite of the importance of behavioral plasticity in biological systems, its implications for social systems have generally gone unexplored.

In investigating this issue, I assumed flexible agents—by choosing the action best suited for the current environment—maximize the probability of surviving today, while rigid agents—by always choosing the same action—become relatively proficient with a particular action. A key assumption is that an agent's current performance (where current survival depends on relative performance) is determined primarily by his current action and secondarily by his proficiency with that action. A central finding is that rigid agents can thrive even when experience and the proficiency it generates counts much less in determining current survival than doing what is right for the moment. It was also found, unsurprisingly, that flexible agents do better when proficiency requires less experience. More interestingly, if there are initially enough rigid agents present, rigid agents will dominate even when proficiency is achieved with a trivial amount of experience. While having current performance be more heavily influenced by one's action than one's proficiency biases the model against rigid agents, a feature of the model that favors rigid agents is having an agent's current survival depend only on current and not past performance because, over all types of environments, flexible agents generate higher average performance in that they always choose the best action—and that is what primarily determines current performance. Relatedly, another feature of the model that would seem to favor rigid agents is the "up-or-out" structure in which agents who perform relatively better advance and the remainder drop out and no longer compete. If an agent who fails to be promoted is not immediately kicked out of the system, the higher average performance of flexible agents may result in them doing better. Indulging in some speculation, this suggests that rigid agents might do relatively better in electoral systems than in corporations in that the former are characterized by an up-or-out structure and arguably put more weight on current than past performance. It is this type of insight that I am striving for and which will help identify how the properties of a social system relate to the behavioral types that thrive in such systems.

In light of the simplicity of our model, it is natural to wonder about the robustness of our results. To begin, one needs to properly pose this issue. If the model were to be generalized to allow for many environments, many actions, and many behavioral types, I certainly would not expect purely rigid agents to thrive. They are an extreme type that perhaps only does well in the extreme model. The pertinent issue is instead whether *relatively* rigid agents do well. Robustness then pertains to whether the identified forces would still be operative in a more general setting and thereby contribute to the determination of which type prevails. I suspect that consistency of action would continue to contribute to rigid agents surviving, the process by which flexible agents lose proficiency would continue to result in nonmonotonic dynamics, and a quicker learning curve for achieving proficiency would continue to promote flexible agents.

While having more environments and actions might seem opportune for flexible agents, let me put forth some speculative arguments as to why this might not necessarily be so. Suppose there are n environments, n actions, and each action is best for some environment. In that they always use the same

action, there are n types of fully rigid agents. With a fully flexible agent using a different action for each environment, there are $n!$ types of fully flexible agents. Let me further suppose that m actions are highly detrimental in any environment other than the one for which they are best while the other $n-m$ actions do reasonably well in all environments. If each possible rule is initially equally represented then a proportion $(n-m)/n$ of purely rigid agents avoids those actions that are highly detrimental in most environments while only a proportion $1/n(n-1)\cdots(n-m+1)$ of fully flexible agents do so. Even if the best rule is one that is fully flexible, one could imagine the subpopulation of purely rigid agents outperforming, at least for a time, the subpopulation of fully flexible agents. Another factor arises if I generalize a step further and suppose that an agent observes the environment with noise. A rigid agent who avoids actions that are detrimental in certain situations may perform better then a flexible agent who includes them in his repertoire. As Heiner (1983) argues, for a bigger behavioral repertoire to be advantageous, an agent must be more skillful at recognizing the environment for which an action is appropriate. This is a consideration that has thus far been ignored in that I have presumed that a flexible agent is able to perfectly deploy his strategy. It is then not so obvious that more environments and more actions are necessarily conducive to flexible agents thriving.

The objective of this particular piece of research was a modest one—to begin to accumulate insight into how a class of competitive selection processes selects among a population of agents who differ in terms of their rigidity. The ultimate goal of this line of research is to generate predictive statements about real-world systems. One class of statements would describe how rigidity varies across levels within a hierarchical social system. For example, should more "yes men" be expected to be found in middle or upper management? In contrast, statements across social systems concern how the overall rigidity of a social system varies with such characteristics as the number of levels and the volatility of the environment. For example, do "yes men" have a smaller presence in flatter organizations? While the model of this paper is far too simple to be a reasonable representation of such a complex system as an electoral system or a corporation, the hope is that later generations of models will be.

Appendix

PROOF OF THEOREM 1:

The proof is comprised of three steps. First, as $t \to \infty$ the proportion of $F1$s is shown to go to zero (Lemma A1). Lemma A1 is then used to show that the proportion of $R1$s is not bounded below 1 (Lemma A2). It follows from Lemmas A1 and A2 that T can be chosen so as to make f_1^t arbitrarily close to zero $\forall t \geq T$ and r_1^T arbitrarily close to one. From this result, it is argued that r_1^t remains arbitrarily close to one $\forall t > T$.

LEMMA A1: $\lim_{t\to\infty} f_1^t = 0$.

PROOF:

Using (1)–(2), it is straightforward to show that $b < 1$ implies that f_1^t/r_1^t is strictly decreasing in t. Since f_1^t/r_1^t is strictly decreasing and f_1^t/r_1^t is bounded below by 0, $\lim_{t\to\infty} f_1^t/r_1^t$ exists and is nonnegative. It follows that:

$$\text{(A1)} \quad \lim_{t\to\infty}[(f_1^{t+1}/r_1^{t+1}) - (f_1^t/r_1^t)] = 0.$$

Using (1)–(2), it is straightforward to show that (A1) is equivalent to:

(A2)

$$\lim_{t\to\infty} \frac{-f_1^t(1-b)}{r_1^t + bf_1^t + 2br_0^t + 2pb(1 - r_1^t - f_1^t - r_0^t)} = 0.$$

Given the denominator is bounded below one and $b < 1$, it follows from (A2) that $\lim_{t\to\infty} f_1^t = 0$.

LEMMA A2: *If $pb > 1/2$ then there does not exist $\bar{r} < 1$ such that $r_1^t < \bar{r}$ for all t.*

PROOF:

Suppose not so that $\exists \bar{r} < 1$ such that $r_1^t < \bar{r}\ \forall t$. Using (1), it is derived that:

$$\text{(A3)}\quad \Delta r_1^t \equiv r_1^{t+1} - r_1^t = r_1^t[(2pb - 1)(1 - r_1^t) - b(2p - 1)f_1^t + 2b(1 - p)r_0^t].$$

By Lemma A1, $\lim_{t\to\infty} - b(2p - 1)f_1^t = 0$. By supposition, $r_1^t < \bar{r}\ \forall t$ which implies, with $pb > 1/2$, that $(2pb - 1)(1 - r_1^t) > (2pb - 1)(1 - \bar{r}) > 0\ \forall t$. From these two properties, it follows that $\exists$ finite T such that $\Delta r_1^t > (2pb - 1)(1 - \bar{r})\ \forall t \geq T$. But if r_1^t is increasing at a rate bounded above zero for all $t \geq T$ then $\lim_{t\to\infty} r_1^t = +\infty$ which is not possible since, by construction, $r_1^t \in [0, 1]\ \forall t$. Given that this contradiction was derived under the supposition that Lemma A2 is false, it is concluded that Lemma A2 is true.

With the aid of Lemmas A1 and A2, I will now prove Theorem 1 by showing that: $\forall \varepsilon > 0\ \exists$ finite T such that $r_1^t \in (1 - \varepsilon, 1)\ \forall t \geq T$. By Lemma A1, $\exists$ finite t' such that $f_1^t < [(2pb - 1)/b(2p - 1)](\varepsilon/2)\ \forall t \geq t'$, where $\varepsilon > 0$ (note that $pb > 1/2$ implies $[(2pb - 1)/b(2p - 1)] > 0$). It is an implication of Lemma A2 that $\forall$ finite t, $\exists$ finite $T \geq t$ such that $r_1^T > 1 - (\varepsilon/2)$. Setting $t = t'$, I conclude that $\exists$ finite T such that $f_1^t < [(2pb - 1)/b(2p - 1)](\varepsilon/2)\ \forall t \geq T$ and $r_1^T > 1 - (\varepsilon/2)$. The next step is to derive sufficient conditions for $\Delta r_1^t > 0$. Since $r_1^t \in (0, 1 - [b(2p - 1)/(2pb - 1)]f_1^t)$ implies $\Delta r_1^t > 0$ and given $f_1^t < [(2pb - 1)/b(2p - 1)](\varepsilon/2)\ \forall t \geq T$, I conclude that if $r_1^t \in (0, 1 - (\varepsilon/2))$ then $\Delta r_1^t > 0\ \forall t \geq T$. Next, a lower bound on Δr_1^t is derived. From (A3) and $p \geq 1/2b > 1/2$, $\Delta r_1^t > -b(2p - 1)r_1^t f_1^t > -bf_1^t$. If $t \geq T$ then $-bf_1^t > -b[(2pb - 1)/b(2p - 1)](\varepsilon/2) = -(2b - 1)(\varepsilon/2) > -\varepsilon/2$. Therefore, $\Delta r_1^t > -\varepsilon/2\ \forall t \geq T$.

To summarize, I have shown that $\forall \varepsilon > 0\ \exists$ finite T such that: (i) $r_1^T > 1 - (\varepsilon/2)$; (ii) if $r_1^t \in (0, 1 - (\varepsilon/2))$ then $\Delta r_1^t > 0\ \forall t \geq T$; and (iii) $\Delta r_1^t > -\varepsilon/2\ \forall t \geq T$. By (i) and (iii), $r_1^{T+1} > 1 - \varepsilon$. Now suppose that $t > T$ and $r_1^t > 1 - \varepsilon$. If $r_1^t \in (1 - \varepsilon, 1 - (\varepsilon/2))$ then, by (ii), $\Delta r_1^t > 0$, which implies $r_1^{t+1} > r_1^t$ and therefore $r_1^{t+1} > 1 - \varepsilon$. If $r_1^t > 1 - (\varepsilon/2)$ then, by (iii), $r_1^{t+1} > r_1^t - (\varepsilon/2)$ and therefore $r_1^{t+1} > 1 - \varepsilon$. I have then shown that $r_1^{T+1} > 1 - \varepsilon$ and if $r_1^t > 1 - \varepsilon$ then $r_1^{t+1} > 1 - \varepsilon$, $t \geq T + 1$. By induction, it is concluded that $r_1^t > 1 - \varepsilon\ \forall t > T$. Since ε was arbitrary, this proves Theorem 1.

PROOF OF THEOREM 2:

Using (1) and (3), after a few steps one can derive:

$$\text{(A4)}\quad \Delta r_1^t + \Delta r_0^t = (1 - r_1^t - r_0^t)[(2pb - 1)r_1^t + (2p(1 - b) - 1)r_0^t].$$

Since $p \leq 1/2b$ then $\Delta r_1^t + \Delta r_0^t < 0$. Given that $r_1^t + r_0^t$ is monotonically decreasing and has a lower bound of zero, $\lim_{t\to\infty} \Delta r_1^t + \Delta r_0^t = 0$. Since $(1 - r_1^t - r_0^t)$ is bounded above zero $\forall t$, it follows from (A4) that $\lim_{t\to\infty} \Delta r_1^t + \Delta r_0^t = 0$ iff $\lim_{t\to\infty} r_1^t = 0$ and $\lim_{t\to\infty} r_0^t = 0$.

PROOF OF THEOREM 3:

It is straightforward to show that:

$$\text{(A5)}\quad \Delta r_1^t = r_1^t(2b - 1)(1 - r_1^t - f_1^t),$$

$$\text{(A6)}\quad \Delta f_1^t = f_1^t[(2b - 1)(1 - r_1^t) - bf_1^t].$$

By the same method used to prove Lemma A1 one can establish that $\lim_{t\to\infty} f_1^t = 0$. By (A5) r_1^t is monotonically increasing. Since r_1^t is bounded above by one, it then has a limit which implies $\lim_{t\to\infty} \Delta r_1^t = 0$. Since $\lim_{t\to\infty} f_1^t = 0$ and r_1^t is increasing then $\lim_{t\to\infty} r_1^t = 1$.

PROOF OF THEOREM 4:

It is straightforward to show that:

$$\text{(A7)}\quad \Delta r_1^t = r_1^t(2b - 1)(1 - r_1^t - f_1^t),$$

$$\text{(A8)}\quad \Delta f_1^t = f_1^t(2b - 1)(1 - r_1^t - f_1^t).$$

Since $r_1^t + f_1^t$ are monotonically increasing and bounded above by one then $r_1^t + f_1^t$ has a limit so that $\lim_{t\to\infty}(\Delta r_1^t + \Delta f_1^t) = 0$. Since $\Delta r_1^t + \Delta f_1^t = (2b-1)(r_1^t + f_1^t)(1 - r_1^t - f_1^t)$ then $\lim_{t\to\infty}(\Delta r_1^t + \Delta f_1^t) = 0$ implies $\lim_{t\to\infty}(r_1^t + f_1^t) = 1$.

PROOF OF THEOREM 5:

The first step is to partition rigid agents into those who have always chosen the action best for the environment and those who have not.

$\hat{r}_i^t \equiv$ proportion of the level t population that are rigid agents endowed with action i and who have always faced a type i environment.

$\bar{r}_i^t \equiv$ proportion of the level t population that are rigid agents endowed with action i and who have previously faced a type j $(\neq i)$ environment.

Note that $r_i^t = \bar{r}_i^t + \hat{r}_i^t$. It is straightforward to derive the equations of motion for $t \geq 2$:

$$\text{(A9)}\quad \bar{r}_1^{t+1} = \bar{r}_1^t[2b(\bar{r}_0^t + \hat{r}_0^t) + 2(1-b)\hat{r}_1^t + \bar{r}_1^t] + (1-b)(\hat{r}_1^t)^2,$$

$$\text{(A10)}\quad \hat{r}_1^{t+1} = \hat{r}_1^t[2b - b\hat{r}_1^t - bf_1^t],$$

$$\text{(A11)}\quad f_1^{t+1} = f_1^t[2b - b\hat{r}_1^t - bf_1^t],$$

$$\text{(A12)}\quad \bar{r}_0^{t+1} = \bar{r}_0^t[2(1-b)(\bar{r}_1^t + \hat{r}_1^t) + 2b\hat{r}_0^t + \bar{r}_0^t] + b(\hat{r}_0^t)^2,$$

$$\text{(A13)}\quad \hat{r}_0^{t+1} = \hat{r}_0^t[2(1-b) - (1-b)\hat{r}_0^t - (1-b)f_0^t],$$

$$\text{(A14)}\quad f_0^{t+1} = f_0^t[2(1-b) - (1-b)\hat{r}_0^t - (1-b)f_0^t].$$

LEMMA A3: $\lim_{t\to\infty} r_0^t = 0$.

PROOF:

Since $\Delta\hat{r}_0^t = \hat{r}_0^t[-(2b-1) - (1-b)\hat{r}_0^t - (1-b)f_0^t] < 0$, it is easy to show that $\lim_{t\to\infty}\hat{r}_0^t = 0$. Using this fact then a little manipulation shows: $\lim_{t\to\infty}\Delta\bar{r}_0^t = \lim_{t\to\infty} -\bar{r}_0^t[1 - \bar{r}_0^t - 2(1-b)(\bar{r}_1^t + \hat{r}_1^t)] \leq 0$. From this it is concluded that $\lim_{t\to\infty}\bar{r}_0^t = 0$ and thus $\lim_{t\to\infty} r_0^t = 0$.

LEMMA A4: $\lim_{t\to\infty}\hat{r}_1^t = [(2b-1)/b][r_1^1/(1-r_0^1)] \equiv \hat{\rho}$.

PROOF:

Since $\hat{r}_1^{t+1} = \hat{r}_1^2\prod_{\tau=2}^t[2b - b\hat{r}_1^\tau - bf_1^\tau]$ and $f_1^{t+1} = f_1^2\prod_{\tau=2}^t[2b - b\hat{r}_1^\tau - bf_1^\tau]$ then $\hat{r}_1^{t+1}/f_1^{t+1} = \hat{r}_1^2/f_1^2$. It is straightforward to derive: $\hat{r}_1^2 = br_1^1(1 + r_0^1)$ and $f_1^2 = b(1 - r_1^1 - r_0^1)(1 + r_0^1)$. It is then obtained that $\hat{r}_1^t/f_1^t = r_1^1/(1 - r_1^1 - r_0^1)\ \forall t \geq 3$. Substituting $[(1 - r_1^1 - r_0^1)/r_1^1]\hat{r}_1^t$ for f_1^t in (A10), the following is derived:

$$\text{(A15)}\quad \hat{r}_1^{t+1} = \hat{r}_1^t[2b - b((1-r_0^1)/r_1^1)\hat{r}_1^t].$$

Using (A15), it is easy to show that $\Delta\hat{r}_1^t \gtreqless 0$ as $\hat{r}_1^t \lesseqgtr \hat{\rho}$. If $\hat{r}_1^t$ has a limit, it must then be $\hat{\rho}$. $\hat{r}_1^t$ converges to $\hat{\rho}$ if $|\hat{r}_1^{t+1} - \hat{\rho}| < |\hat{r}_1^t - \hat{\rho}|$. This will only be shown for $\hat{r}_1^t < \hat{\rho}$, as the proof is analogous for $\hat{r}_1^t > \hat{\rho}$. Given $\hat{r}_1^t < \hat{\rho}$ and therefore $\hat{r}_1^{t+1} > \hat{r}_1^t$, if $\hat{r}_1^{t+1} < \hat{\rho}$ then I am done. Suppose instead that $\hat{r}_1^{t+1} > \hat{\rho}$. Using (A15) and substituting $((2b-1)/\hat{\rho})$ for $b((1-r_0^1)/r_1^1)$, one derives that $|\hat{r}_1^{t+1} - \hat{\rho}| < |\hat{r}_1^t - \hat{\rho}|$ if and only if:

$$\text{(A16)}\quad \varphi(\hat{r}_1^t) \equiv 2\hat{\rho} - (2b+1)\hat{r}_1^t + (2b-1)(1/\hat{\rho})(\hat{r}_1^t)^2 > 0.$$

Since $\varphi'(\hat{\rho}) < 0$, and $\varphi''(\cdot) > 0\ \forall\hat{r}_1^t$ then $\varphi(\hat{r}_1^t) > \varphi(\hat{\rho})\ \forall\hat{r}_1^t < \hat{\rho}$. Given $\varphi(\hat{\rho}) = 0$, it can be concluded that (A16) is true. Hence, $\lim_{t\to\infty}\hat{r}_1^t = \hat{\rho}$.

LEMMA A5: $\lim_{t\to\infty}\bar{r}_1^t = (1/2)\{1 - 2(1-b)\hat{\rho} - [1 - 4(1-b)\hat{\rho} - 4b(1-b)\hat{\rho}^2]^{1/2}\} \equiv \bar{\rho}$.

PROOF:

Since $\lim_{t\to\infty} r_0^t = 0$ and $\lim_{t\to\infty} \hat{r}_1^t = \hat{p}$, it follows that:

$$\text{(A17)}\quad \lim_{t\to\infty} \left| (\partial \bar{r}_1^{t+1}/\partial \bar{r}_1^t) - [2(1-b)\hat{p} + 2\bar{r}_1^t] \right| = 0.$$

Thus, as $t \to \infty$, $(\partial \bar{r}_1^{t+1}/\partial \bar{r}_1^t) < 1$ iff $\bar{r}_1^t < (1/2) - (1-b)\hat{p}$. It follows that if $\bar{r}_1^t < (1/2) - (1-b)\hat{p}$ as $t \to \infty$ then $\bar{r}_1^t$ converges. Furthermore, it can be shown that, as $t \to \infty$, $\Delta \bar{r}_1^t \gtreqless 0$ as $\lim_{t\to\infty} \bar{r}_1^t \gtreqless \bar{p}$. Hence, if $\bar{r}_1^t$ converges it must converge to $\bar{p}$. Given that $\bar{p} < [(1/2) - (1-b)\hat{p}]$ then, as $t \to \infty$, $\Delta \bar{r}_1^t < 0$ when $\bar{r}_1^t > (1/2)[1 - 2(1-b)\hat{p}]$. Therefore, $\exists\, T$ such that $\forall t \geq T$ if $\bar{r}_1^t > (1/2)[1 - 2(1-b)\hat{p}]$ then $\Delta \bar{r}_1^t < 0$. This implies $\exists\, T' > T$ such that $\bar{r}_1^{T'} < [(1/2) - (1-b)\hat{p}]$. Since then $|(\partial \bar{r}_1^{t+1}/\partial \bar{r}_1^t)| < 1\ \forall t \geq T'$, it follows that $\bar{r}_1^t$ converges to $\bar{p}$.

I now want to show that $r_1^1 + r_0^1 > \hat{p} + \bar{p}$. Substituting $(1 - r_0^1)[b/(2b-1)]\hat{p}$ for r_1^1 and using the definition of $\bar{p}$, $r_1^1 + r_0^1 > \hat{p} + \bar{p}$ is equivalent to:

$$\text{(A18)}\quad (1 - r_0^1)[b/(2b-1)]\hat{p} + r_0^1 > \hat{p} + (1/2)\{1 - 2(1-b)\hat{p} - [1 - 4(1-b)\hat{p} - 4b(1-b)\hat{p}^2]^{1/2}\}.$$

The left-hand side can be rearranged to: $[b/(2b-1)]\hat{p} + [1 - (b/(2b-1))\hat{p}]r_0^1$; and it is straightforward to show that $[1 - (b/(2b-1))\hat{p}]r_0^1 > 0$. Hence, (A18) holds if:

$$\text{(A19)}\quad [b/(2b-1)]\hat{p} > \hat{p} + (1/2)\{1 - 2(1-b)\hat{p} - [1 - 4(1-b)\hat{p} - 4b(1-b)\hat{p}^2]^{1/2}\}.$$

Working through a few algebraic steps shows that (A19) is equivalent to: $(2b-1)/b > \hat{p}$, which is indeed true.

PROOF OF THEOREM 6:

Since an asymptotic attractor must be a rest point, the set of rest points is first derived and then their asymptotic stability is evaluated.

PROPOSITION A1: *(r_1, f_1, r_0, f_0) is a rest point if and only if $(r_1, f_1, r_0, f_0) \in \{(1, 0, 0, 0), (0, 0, 1, 0), (0, b, 0, 1-b), ((1-b)/b, (2b-1)^2/b^2, 0, (1-b)(2b-1)/b^2)\}$.*

PROOF:

The proof is by construction. Substituting $1 - r_1^t - f_1^t - f_0^t$ for r_0^t in (10), (11), and (13), the dynamical system can be represented by:

$$\text{(A20)}\quad \Delta r_1^t = r_1^t[(2b-1)(1-r_1^t) - bf_1^t],$$

$$\text{(A21)}\quad \Delta f_1^t = f_1^t[(2b-1) - br_1^t - bf_1^t] + f_0^t[2b(1-r_1^t) - 2bf_1^t - bf_0^t],$$

$$\text{(A22)}\quad \Delta f_0^t = f_0^t[-b + (1-b)r_1^t + (1-b)f_1^t] + f_1^t[2(1-b)r_1^t + (1-b)f_1^t].$$

$(r_1, f_1, 1 - r_1 - f_1 - f_0, f_0)$ is a rest point if and only if (A20)–(A22) equal zero. If $r_0 = 1$ then $r_1 = f_1 = f_0 = 0$. Since (A20)–(A22) equal zero, (0, 0, 1, 0) is a rest point. Given $\Delta r_0^t = -r_0^t[(2b-1)(r_1^t + f_1^t) + bf_0^t]$ then $\Delta r_0^t < 0$ when $r_1^t < 1$. Hence, if (r_1, f_1, r_0, f_0) is a rest point and $r_0 < 1$ then $r_0 = 0$.

Since $r_0 = 0$ at all other rest points, a rest point is defined by (r_1, f_1, f_0) such that $\Delta r_1^t = 0$, $\Delta f_1^t = 0$, $\Delta f_0^t = 0$, and $r_1 + f_1 + f_0 = 1$. Substitute $1 - r_1 - f_1$ for f_0 in (A20)–(A21). Simplifying, $(r_1, f_1, 0, 1 - r_1 - f_1)$ is a rest point if and only if (r_1, f_1) satisfies:

$$\text{(A23)}\quad r_1[(2b-1)(1-r_1) - bf_1] = 0,$$

$$\text{(A24)}\quad b - f_1(1 - br_1) - br_1(2 - r_1) = 0.$$

If $r_1 = 0$ then (A23) is satisfied while (A24) is satisfied by $f_1 = b$. Hence, $(0, b, 0, 1 - b)$ is a rest point. Now consider $r_1 = 1$. Since then $f_1 = 0$, (A23)–(A24) are satisfied so that $(1, 0, 0, 0)$ is a rest point. The final case to consider are rest points for which $r_1 \in (0, 1)$. Solving (A23)–(A24), one derives $f_1 = ((2b - 1)/b)^2$ and $r_1 = (1 - b)/b$. The unique rest point for which $r_1 \in (0, 1)$ is then $((1 - b)/b, (2b - 1)^2/b^2, 0, (1 - b)(2b - 1)/b^2)$.

To prove that $(0, 0, 1, 0)$ is not an asymptotic attractor, (12) is used to derive $\Delta r_0^t = -r_0^t[(2b - 1)(r_1^t + f_1^t) + bf_0^t]$. Since $\Delta r_0^t < 0$ $\forall r_0^t \in (0, 1)$, it follows that $\lim_{t\to\infty} r_0^t = 0 < 1$.

To prove that $((1 - b)/b, (2b - 1)^2/b^2, 0, (1 - b)(2b - 1)/b^2)$ is not an asymptotic attractor, it will be shown that in any neighborhood of this rest point, there exists a set of points such that the system does not converge to this rest point. Consider an initial state such that: $r_1^1 > (1 - b)/b$, $r_0^1 = 0$, and

$$\text{(A25)} \quad [(2b - 1)/b](1 - r_1^1) > f_1^1 > b(1 - r_1^1)^2/(1 - br_1^1).$$

Note that the expressions on either side of f_1^1 equal $(2b - 1)^2/b^2$ when $r_1^1 = (1 - b)/b$. Therefore, by choosing r_1^1 sufficiently close to $(1 - b)/b$ and having f_1^1 satisfy (A25), this state can be made arbitrarily close to the rest point. I want to show that $\Delta r_1^t > 0$ $\forall t \geq 1$. Since r_1^1 exceeds its value at the rest point, it would follow that the system diverges.

From (A20), one knows that $\Delta r_1^1 > 0$ iff $r_1^1 < 1 - [b/(2b - 1)]f_1^1$. Rearranging the left-hand side of (A25), one finds that $r_1^1 < 1 - [b/(2b - 1)]f_1^1$. Hence, $\Delta r_1^1 > 0$. When $r_0^1 = 0$, one knows that $\Delta f_1^1 > 0$ iff $f_1^1 > b(1 - r_1^1)^2/(1 - br_1^1)$. Thus, by the right-hand-side inequality in (A25), it is concluded that $\Delta f_1^1 < 0$. I next want to show: if $\Delta r_1^t > 0$ and $\Delta f_1^t < 0$ then $\Delta r_1^{t+1} > 0$ and $\Delta f_1^{t+1} < 0$. If $\Delta r_1^t > 0$ then $r_1^t < 1 - [b/(2b - 1)]f_1^t$ so that $r_1^t = 1 - [b/(2b - 1)]f_1^t - \varepsilon$ for some $\varepsilon > 0$. Using (10), it follows that: $r_1^{t+1} = 1 - [b/(2b - 1)]f_1^t - \varepsilon[2(1 - b) + bf_1^t + \varepsilon]$; and therefore $r_1^{t+1} < 1 - [b/(2b - 1)]f_1^t$. Since, by supposition, $\Delta f_1^t < 0$ then $f_1^{t+1} < f_1^t$. It follows from $r_1^{t+1} < 1 - [b/(2b - 1)]f_1^t$ that $r_1^{t+1} < 1 - [b/(2b - 1)]f_1^{t+1}$. It is concluded that $\Delta r_1^{t+1} > 0$. Now I turn to show that $\Delta f_1^{t+1} < 0$. Since $r_0^t = 0$ and $\Delta f_1^t < 0$ then $f_1^t = [b(1 - r_1^t)^2/(1 - br_1^t)] + \varepsilon$ for some $\varepsilon > 0$. From (11), $f_1^{t+1} = [b(1 - r_1^t)^2/(1 - br_1^t)] + br_1^t\varepsilon$ so that $f_1^{t+1} > b(1 - r_1^t)^2/(1 - br_1^t)$. Since $\Delta r_1^t > 0$ then $r_1^{t+1} > r_1^t$. Given that $b(1 - r_1^t)^2/(1 - br_1^t)$ is decreasing in r_1^t, it follows that $f_1^{t+1} > b(1 - r_1^{t+1})^2/(1 - br_1^{t+1})$ and thus $\Delta f_1^{t+1} < 0$.

It has been shown that: (i) $\Delta r_1^1 > 0$ and $\Delta f_1^1 < 0$; and (ii) if $\Delta r_1^t > 0$ and $\Delta f_1^t < 0$ then $\Delta r_1^{t+1} > 0$ and $\Delta f_1^{t+1} < 0$. By induction, $\Delta r_1^t > 0$ and $\Delta f_1^t < 0$ $\forall t \geq 1$. It is inferred from $r_1^1 > (1 - b)/b$ and $\Delta r_1^t > 0$ $\forall t \geq 1$ that $\lim_{t\to\infty} r_1^t > (1 - b)/b$. This establishes that $((1 - b)/b, (2b - 1)^2/b^2, 0, (1 - b)(2b - 1)/b^2)$ is not an asymptotic attractor.

The final step is to show that $(1, 0, 0, 0)$ and $(0, b, 0, 1 - b)$ are asymptotic attractors. Substituting $1 - r_1^t - f_1^t - f_0^t$ for r_0^t in (10), (11), and (13) and defining $\Pi(f_1^t, r_0^t, f_0^t)$ to be the matrix of first derivatives, it is straightforward to derive (A26) below.

Let $\Lambda(r_1^t, f_1^t, f_0^t) \equiv \max\{\lambda_1, \ldots, \lambda_k\}$ where $\lambda_1, \ldots, \lambda_k$ are the eigenvalues to $\Pi(r_1^t, f_1^t, f_0^t)$. By Theorem 4.10 in Kelley and Peterson (1991), if $\Lambda(r_1, f_1, f_0) < 1$ then (r_1, f_1, f_0) is an asymptotic attractor. Using (A26), one can show that $\Lambda(1, 0, 0) = \max\{2(1 - b), b, 1 - b\} < 1$ and $\Lambda(0, b, 1 - b) = \max\{b(2 -$

$$\text{(A26)} \quad \Pi(r_1^t, f_1^t, f_0^t) = \begin{bmatrix} 2b - 2(2b-1)r_1^t - bf_1^t & -br_1^t & 0 \\ -b(f_1^t + 2f_0^t) & b(2 - r_1^t - 2f_1^t - 2f_0^t) & 2b(1 - r_1^t - f_1^t - f_0^t) \\ (1-b)(2f_1^t + f_0^t) & (1-b)(2r_1^t + 2f_1^t + f_0^t) & (1-b)(1 + r_1^t + f_1^t) \end{bmatrix}.$$

$b), 0, 1 - b^2\} < 1$. Therefore, $(r_1, f_1, r_0, f_0) = (1, 0, 0, 0)$ and $(r_1, f_1, r_0, f_0) = (0, b, 0, 1 - b)$ are asymptotic attractors.

REFERENCES

Ben-Porath, Elchanan; Dekel, Eddie and Rustichini, Aldo. "On the Relationship between Mutation Rates and Growth Rates in a Changing Environment." *Games and Economic Behavior*, October 1993, *5*(4), pp. 576–603.

Bickerton, Derek. *Language and species*. Chicago: University of Chicago Press, 1990.

Boylan, Richard T. "Laws of Large Numbers for Dynamical Systems with Randomly Matched Individuals." *Journal of Economic Theory*, August 1992, *57*(2), pp. 473–504.

Burton, Steven J. *Judging in good faith*. Cambridge: Cambridge University Press, 1992.

Guth, William D. and Tagiuri, Renato. "Personal Values and Corporate Strategy." *Harvard Business Review*, September–October 1965, *43*(5), pp. 123–32.

Haltiwanger, John and Waldman, Michael. "Responders versus Non-Responders: A New Perspective on Heterogeneity." *Economic Journal*, September 1991, *101*(408), pp. 1085–102.

Harrington, Joseph E., Jr. "The Social Selection of Principled and Expedient Agents." Department of Economics Working Paper No. 324, Johns Hopkins University, March 1994.

———. "Social Learning and Rigid Behavior." Mimeo, Johns Hopkins University, February 1996.

Hayes, Robert H. and Pisano, Gary P. "Beyond World Class: The New Manufacturing Strategy." *Harvard Business Review*, January–February 1994, *72*(1), pp. 77–86.

Heiner, Ronald A. "The Origin of Predictable Behavior." *American Economic Review*, September 1983, *73*(4), pp. 560–95.

Jonassen, David H. and Grabowski, Barbara L. *Handbook of individual differences, learning and instruction*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.

Kelley, Walter G. and Peterson, Allan C. *Difference equations*. Boston: Harcourt Brace Jovanovich, 1991.

Kydland, Finn E. and Prescott, Edward C. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, June 1977, *85*(3), pp. 473–91.

Luchins, Abraham S. and Luchins, Edith Hirsch. *Rigidity of behavior*. Eugene, OR: University of Oregon Books, 1959.

Manchester, William. *Alone 1932–1940*. Boston: Little, Brown & Co., 1988.

Maranell, Gary M. "The Evaluation of Presidents: An Extension of the Schlesinger Polls." *Journal of American History*, June 1970, *57*(1), pp. 104–13.

Rosen, Sherwin. "Prizes and Incentives in Elimination Tournaments." *American Economic Review*, September 1986, *76*(4), pp. 701–15.

Rosenthal, Robert W. "Rules of Thumb in Games." *Journal of Economic Behavior and Organization*, September 1993, *22*(1), pp. 1–13.

Sah, Raaj K. and Stiglitz, Joseph E. "The Quality of Managers." *Quarterly Journal of Economics*, February 1991, *106*(1), pp. 289–95.

Schlesinger, Joseph A. *Ambition and politics: Political careers in the United States*. Chicago: Rand McNally, 1966.

———. *Political parties and the winning of office*. Ann Arbor, MI: University of Michigan Press, 1991.

Scott, William A. "Flexibility, Rigidity, and Adaptation: Toward Clarification of Concepts," in O. J. Harvey, ed., *Experience, structure, and adaptability*. New York: Springer, 1966, pp. 369–400.

Sloan, Alfred P. *My years with General Motors*. New York: Macfadden-Bartell, 1965.

Stahl, Dale O. "Evolution of Smart$_n$ Players." *Games and Economic Behavior*, October 1993, *5*(4), pp. 604–17.

Walton, Sam and Huey, John. *Sam Walton: Made in America*. New York: Doubleday, 1992.

Wildavsky, Aaron. "The Goldwater Phenomenon: Purists, Politicians, and the Two-Party System." *Review of Politics*, July 1965, *27*(3), pp. 386–413.

Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates

By PETER A. DIAMOND*

Using the Mirrlees optimal income tax model with quasi-linear preferences, the paper examines conditions for marginal tax rates to be rising at high income levels and declining in an interval containing the modal skill. It examines conditions for the marginal tax rate to be higher at a low skill level than at the high skill level with the same density—an argument only holding for skill levels above a cutoff where resources of a worker are marginally of the same value as resources of the government. Data on earnings rates are presented. (*JEL* H21)

The trade-off between efficiency and income distribution plays a central role in analyzing tax policy.[1] The modern framework for analyzing this trade-off using nonlinear income taxes was created in James A. Mirrlees (1971). While this formulation crystallized a presentation of the income tax problem and derived some of the properties of optimal income taxation, the implications for policy have been somewhat limited. For example, the *Financial Times* (September 11, 1995 p. 24) has summarized the policy impact of the optimal income tax literature as: "A few general principles none the less gained the status of received wisdom, for example that marginal tax rates should be constant and modest over most of the income range, but zero at the top and bottom." The public finance community has recognized that the results deriving zero marginal tax rates at the top and the bottom of the income distribution are of little or no relevance for policy. This paper argues that the case for nonconstant and high marginal tax rates in the Mirrlees model is considerably stronger than has been realized. The technical contribution of this paper is very modest, being primarily a rearrangement of terms in the standard first-order condition for optimal income taxation. This rearrangement leads to a different way of approaching the combinations of assumptions that will sign the change in marginal tax rates with income level. In addition, by concentrating on the case where there is a zero income derivative of labor supply, the intuition behind the first-order condition becomes clearer.

Section I reviews some of the previous literature. Section II presents the optimal income tax problem. Section III examines conditions for marginal tax rates to be rising at income levels above the modal skill level. Section IV examines the level of marginal tax rates on very high incomes. Section V examines conditions for the marginal tax rate to be declining and to be higher at a skill level below the modal skill than at the skill level above the mode with the same density of skills. These arguments apply for skill levels above a cutoff level, where resources are of the same value in the hands of the government and in the hands of a worker with the cutoff skill level. Section VI considers another example. Section VII looks at data on earnings rates to suggest the relevance of alternative empirical assumptions on the distribution of skills. Some closing remarks are in Section VIII.

* Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139. I am grateful to Jon Gruber for suggesting the inclusion of data in this paper and providing me with the figures presented below, and to Marcus Berliant, Svetlana Danilkina, Jim Mirrlees, Eytan Sheshinski, Jay Wilson, and anonymous reviewers for helpful suggestions. I am also grateful to the National Science Foundation for research support under Grant SBR-9307876.

[1] If part of the population has potential income below per capita government expenditures, then it is impossible to finance government spending without some distorting taxes. This paper assumes that concern for the income of the poor is large enough that there is a positive transfer to those with zero income.

I. Review of the Literature

Formal results about the optimal income tax are fairly limited. (For recent expositions, see Matti Tuomala, 1990 Ch. 6 or Gareth D. Myles, 1995 Ch. 5.) Assuming that labor supply can be continuously adjusted, there is no gain from having marginal tax rates above 100 percent since no one will have such a tax at the margin. That is, the same outcome can be achieved with taxes no greater than 100 percent. It is usually assumed that preferences are such that consumption is an increasing function of the wage. Then, earnings will be nondecreasing in skill. It then follows that the optimal tax structure has nonnegative marginal rates (Mirrlees, 1971) and positive rates in the interior of the income distribution (Jesus K. Seade, 1982).

Assuming that there is a finite maximum to the skill distribution, the marginal tax rate should be zero at the income level of the top skill (Efraim Sadka, 1976; Seade, 1977). The argument for this result is quite intuitive. Assume this were not the case, then, extending the tax function to higher incomes with a zero tax rate would lead the top earner to work more, raising social welfare without losing any tax revenue. However, this condition need not convey information about optimal taxes over any significant region of incomes—the optimal rates need not approach zero until very close to the top. This point has been made by the numerical calculations in Tuomala (1984).

At the bottom of the skill distribution, in the presence of optimal taxes, there may or may not be an atom of individuals doing no work. If everyone works, then the argument for a zero marginal tax rate carries over (Seade, 1977). However, if there is an atom of nonworkers, the optimal tax has a positive marginal tax rate at the level where earnings begin (Udo Ebert, 1992). This latter case seems empirically more relevant.

In addition to these analytical results, presentation of the first-order condition for optimal taxes has generally been accompanied by observations on the factors leading to high or low rates, ceteris paribus. Considerable effort has gone into simulations, starting with that by Mirrlees. In his simulation Mirrlees assumed a utility function $u = \log[x] + \log[1 - y]$, where x is consumption and y is labor supply (in percentage terms), a social evaluation $G(u) = -\exp(-bu)/b$ $(b > 0)$, and a lognormal distribution of skills. He concluded (p. 206) that "Perhaps the most striking feature of the results is the closeness to linearity of the tax schedules." As seen in the survey in Tuomala (1990), similar results followed with some other simulations, but some simulations have shown other patterns, including a significantly inverse-U shaped pattern (e.g., Ravi Kanbur and Tuomala, 1994). As will be clarified below, simulation results are sensitive to both the utility function and the family of distributions of skills assumed, opening up the possibility of different conclusions.

II. Optimal Income Tax Problem

The Mirrlees optimal income tax problem is the maximization of the integral over the population of a concave function of individual utilities, subject to an aggregate budget constraint and subject to the constraint that individuals optimize in their choice of labor supply given the relationship between work and after-tax income. The only difference across individuals in the model is a difference in skills, with an individual of skill n having a marginal product equal to n. The model is a one-period model with only labor income. It is assumed that the government can observe income received but not hours worked or skill. Denoting consumption of someone with skill n by $x(n)$, labor (in percentage terms) by $y(n)$, and the concave utility function by $u(x, y)$, the social objective function can be stated as

$$(1) \quad \int_{n_0}^{n_1} G\{u[x(n), y(n)]\} f(n)\, dn,$$

where $G(u)$ is an increasing and strictly concave function of utility, with G independent of n, and the distribution of skills is written $F(n)$, with density $f(n)$. It is assumed that the distribution of skills is single-peaked, with mode at n_m. The density is assumed to be positive and continuous between the bottom and the top skill levels, n_0 and n_1.

The resource constraint on this maximization can be stated in terms of output—that aggregate consumption be less than aggregate production minus government expenditures, E:

$$(2) \qquad \int_{n_0}^{n_1} x(n)f(n)\,dn \leqq \int_{n_0}^{n_1} ny(n)f(n)\,dn - E.$$

This constraint can be stated alternatively in terms of taxes. Denoting taxes as a function of earnings as $T[ny(n)]$, consumption equals the difference between earnings and taxes, $x(n) = ny(n) - T[ny(n)]$. In this case, the government budget constraint is that taxes cover government expenditures:

$$(3) \qquad \int_{n_0}^{n_1} T[ny(n)]f(n)\,dn \geqq E.$$

That the resource constraint can be stated equivalently in terms of government budget balance or in terms of aggregate supply and demand is a consequence of Walras Law.

In addition to the resource constraint, there is an incentive compatibility constraint. The government observes earnings, not hours worked or skill. Thus the government is restricted to setting taxes as a function only of earnings. The incentive compatibility constraint is that the selected labor supply, $y(n)$, maximizes utility, given the tax function, $u\{ny(n) - T[ny(n)], y(n)\}$. The relevant part of the tax function is just the part that is selected by someone—taxes can be set arbitrarily high at earnings levels that no one chooses with the optimal tax structure. Thus the incentive compatibility constraint can be stated in the familiar form that a worker with skill n does not prefer to imitate the earnings of a worker with a different skill level:

$$(4) \qquad u\{ny(n) - T[ny(n)], y(n)\} \geqq u\{n'y(n') - T[n'y(n')], n'y(n')/n\}$$

for all n and n'.

That is, someone of skill n would have to work n'/n times as much as someone with skill n' in order to have the same earnings level.

This paper will concentrate on the special case where there are no income effects on labor supply.[2] That is, it is assumed that utility is linear in consumption (referred to as quasi-linear):

$$(5) \qquad u(x, y) = x + v(1 - y) = ny - T(ny) + v(1 - y),$$

where v is assumed to be strictly concave.[3] This assumption seems appropriate at very high income levels, since people at the top of the income distribution are likely to leave large estates—with a linear utility of bequests, neither consumption nor earnings vary with the exact level of estate. (The receipt of such bequests is not part of the model.) In addition, this assumption removes a source of considerable complication in tax analysis. In the presence of distorting taxes, income effects imply that lump-sum taxes have efficiency effects since they change distorted labor supply decisions.

This problem has some complexity in the derivation of the first-order condition for an optimal tax function, but is familiar from a number of mechanism design problems.[4] The simplest way to proceed is to replace the incentive compatibility conditions, (4), with the first-order condition for individual choice, which, from (5), can be written:

$$(6) \qquad v'[1 - y(n)] = n\{1 - T'[ny(n)]\},$$

where T' is the marginal tax rate. For later use, it is convenient to note that for the quasi-linear utility function the elasticity of labor supply

[2] For a discussion of this case with a constant elasticity of labor supply, see Anthony B. Atkinson (1990). The complementary case where utility is linear in leisure has been studied; see Stefan Lollivier and Jean-Charles Rochet (1983) and John A. Weymark (1987).

[3] I assume that $G'[v(1)]$ is infinite, so that someone doing no work is given positive consumption and the non-negativity constraint on consumption can be ignored.

[4] For an exposition of the mechanism design problem, see Drew Fudenberg and Jean Tirole, 1991 Ch. 7.

evaluated at the chosen labor supply of a worker of skill n, $e(n)$, satisfies

$$(7)\quad e(n) = -v'[1 - y(n)] \div \{y(n)v''[1 - y(n)]\}.$$

Since the wage equals the skill level, this is the elasticity with respect to the wage, evaluated at the labor supply level that is chosen by someone with skill n.

More complicated than deriving the first-order condition is the problem of checking when the first-order condition does indeed characterize an optimum. The complication comes from the need to check that individual labor supplies satisfying the first-order conditions are globally optimal choices, and not just the solution to a first-order condition. This problem arises since the budget set is not convex when marginal tax rates are declining over some income levels. For any particular economy, one can check whether individual labor supplies are optimal. This issue raises the possibility that with the optimal tax, the distribution of skills results in a distribution of incomes that either has bunching at some income level (an atom of workers choosing the same income level) or a gap in the distribution of incomes. (Bunching at zero income or a gap between zero and the lowest positive income are not issues for the interpretation of the optimal tax structure below.) I do not explore this issue for this particular class of preferences, but proceed with analysis of the first-order condition; the analysis holds where the equilibrium distribution of incomes has no bunching and no gap, since generically the equation is a necessary condition for the optimal tax where this is true.

The first-order condition for the optimal tax can be calculated by specializing the condition in Mirrlees (1971) for quasi-linear preferences or deriving it directly, as is done in the Appendix. As usually written, the condition is:

$$(8)\quad p(n - v')f = [(v' - yv'')/n] \times \left[\int_n^{n_1} (p - G')\, dF\right],$$

where p is the Lagrange multiplier on the government's budget constraint and the functions v and G are evaluated at the appropriate labor supplies and consumption levels.

It is convenient to use the elasticity of labor supply and the marginal condition for individual choice, (6), to rewrite (8) as

$$(9)\quad T'/(1 - T') = [(e^{-1} + 1)/n] \times \left[\int_n^{n_1} (p - G')\, dF\right] \Big/ [pf].$$

Multiplying and dividing (9) by $(1 - F)$ to turn the integral into an average term, (9) can be rewritten as:

$$(10)\quad T'/(1 - T') = A(n)B(n)C(n)$$

where $A(n) = e^{-1}(n) + 1$;

$$B(n) = \int_n^{n_1} (p - G')\, dF / \{p[1 - F(n)]\};$$

$$C(n) = [1 - F(n)]/[nf(n)].$$

The analysis below examines these three functions, $A(n)$, $B(n)$, and $C(n)$, under alternative assumptions on the functions $e(n)$, $f(n)$, and $G(u)$.

The absence of income effects allows an intuitive grasp of the factors that determine the optimal tax structure. Increasing the marginal tax rate affecting some skill level involves an increase in the deadweight burden for people at this skill level. Thus, the optimal marginal tax rate at some income level depends on the elasticity of labor supply at that income level, since this is important for marginal distortions. Increasing the marginal tax rate also transfers income from all individuals with higher skills to the government, without changing the distortions of their labor supplies. The weights of these two elements depend on the ratio of individuals with skills above this level to individuals with skills at this level and on the level of skill which links the tax on hours to the tax on income. This intuition is displayed in equation (10), where the first-order condition

the optimal income tax is written as a product of these three terms.

The same approach to signing the change in marginal tax rates can be used with the assumption of additive preferences, $u_{xy} = 0$, without the further assumption of quasi-linear preferences. The mathematical conditions for signing the change in the marginal tax rate are similar, although the economic interpretation of the conditions is more complex. $A(n)$ no longer depends only on the compensated elasticity of labor supply, but also has a term involving the second derivative of the utility-of-consumption, which is no longer assumed to be zero. Thus different economic assumptions are needed to sign the mathematical expressions.

III. Increasing Marginal Tax Rates

I turn now to analysis of (10), the first-order condition for the optimal income tax in the presence of quasi-linear preferences. In general, the variation in the elasticity of labor supply with skill will depend on the tax function, since taxes will affect the level of labor supplied and the elasticity varies with the quantity of labor supplied. One obvious exception, making for simpler analysis, is that of a constant elasticity of labor supply. In this case the utility of leisure satisfies $v(1 - y) = c\{1 - [1 - (1 - y)]^k\} = c(1 - y^k)$ for some constants c and k.

LEMMA A: *If* $v(1 - y) = c\{1 - [1 - (1 - y)]^k\} = c(1 - y^k)$, *then* $A(n)$ *is a constant.*

With quasi-linear preferences, a uniform transfer from the government to all workers has no effect on labor supply, and so no extra impact on the government budget. The welfare impact of such a transfer is the average of G' over the entire population. Thus one can conclude that the Lagrangian on the government budget constraint, p, is equal to the average of G':

$$(11) \qquad p = \int_{n_0}^{n_1} G'(n)f(n)\,dn.$$

Thus, $B(n_0)$ is equal to zero.

Given the incentive compatibility constraint, utility must be nondecreasing in skill and increasing where earnings are positive, since a worker can always have the same consumption as a worker with lower skill while doing less work, provided the level of work is positive. That is, above the skills at which there is no work, utility is increasing in n. With G a concave function, G' is then decreasing in n. Since $B(n)$ is the average of $[p - G']$ from the level n to the top of the skill distribution, $B(n)$ is increasing in n.[5]

Since p is equal to the average of G' and G' is nonincreasing, there is a critical value of n, denoted n_C, at which G' is equal to p:

$$(12) \qquad G'[u(n_C)] = p.$$

If n_C occurs at a level of skill where there is positive work, then n_C is unique; otherwise n_C is set equal to the highest skill at which there is no work. The level of n_C is endogenous, varying with both the structure of the economy and the nature of the social welfare function. To simplify the statement of results, analysis is restricted to economies where this critical level is below the modal level of skill:

$$(13) \qquad n_C < n_m.$$

This seems like the more interesting case, assuming that the mode of skills is near the median and the government would like to redistribute toward a fraction of the labor force well below one-half.

I note that $[1 - F(n)]B(n)$ is increasing in n up to n_C and then decreasing in n. These results are summarized as:

LEMMA B: $B(n)$ *is increasing in* n. $[1 - F(n)]B(n)$ *decreases in* n *for* $n > n_C$.

I turn now to the shape of the distribution of skills. Given the assumption of a single-

[5] Formally, differentiating $B(n)$, the derivative has the same sign as the average of $(p - G')$ from n to n_1 minus the value $[p - G'(n)]$. Since G' is decreasing in n, this difference is positive.

peaked density of skills, $nC(n)$ is decreasing in n for n below the modal level, n_m. For values of n above the modal level, the shape of $C(n)$ depends on the family of distributions assumed for skills. With a Pareto distribution above the modal skill level, (i.e., the density is proportional to $1/n^{1+a}$ for $a > 0$), then $C(n)$ is a constant above the modal skill level.

LEMMA C: *For $n < n_m$, $nC(n)$ is decreasing in n. For $n > n_m$, $C(n)$ is constant if $F(n)$ is the Pareto distribution above n_m.*

One can now put these lemmas together to identify sufficient conditions for marginal tax rates to be increasing with income for incomes above the modal level. Where all three of $A(n)$, $B(n)$, and $C(n)$ are nondecreasing and at least one is increasing, then marginal tax rates are increasing.

PROPOSITION 1: *Marginal tax rates are increasing above the modal skill if, above this skill, the elasticity of labor supply is constant and the distribution of skills is Pareto.*

With the conditions in Proposition 1, $A(n)$ and $C(n)$ are constants, so that $T'/(1-T')$ varies with n as $B(n)$ varies with n. With $B(n)$ increasing, so too is T'. The result carries over if the elasticity of labor supply falls with skill at the equilibrium labor supplies. Similarly, it is sufficient to have a distribution of skills such that $[1 - F(n)]/[nf(n)]$ is increasing. Moreover, the result of rising tax rates will hold for part of the skill distribution (above the mode) if the conditions are met for that part; one does not need conditions on the entire distribution.

IV. Asymptotic Marginal Tax Rates

With a known finite top to the distribution of skills, the optimal marginal tax rate is zero at the top of the income distribution. As noted in the review of the literature and is clear from the argument behind Proposition 1, this need not imply that rates approach zero until very close to the top. Thus it is natural to consider the case of an unbounded distribution of skills and to consider the behavior of the optimal marginal tax rate as skills rise without limit.

In addition to assumptions on the distribution of skills and the elasticity of labor supply, the shape of the social welfare of individual utility, $G(u)$, needs to be examined. One possibility is that the marginal welfare weight of consumption of those at the top tends to zero as skill rises without limit[6]. For example, this is the case in the example in Mirrlees (1971) where $G = -\exp(-bu)/b$ $(b > 0)$ and $u = a\log(x) + \log(1-y)$. Similarly, it is the case in Martin Feldstein's (1985) study of social security, where $G = u$ and $u = \log(x)$. If G' goes to zero as n rises without limit, then $B(n)$ goes to 1. Alternatively, one might assume that G' has a positive lower bound which is approached as n rises without limit. For example, Atkinson (1990) considers the case of a "charitable Conservative" position, where the marginal welfare weight of consumption takes on two values—a high one for "poor" people and a low one for "nonpoor" people. I denote by g the ratio of the lower bound on G' to the Lagrangian on the government budget constraint, which is equal to the average of G' in the entire population. Thus, $B(n)$ converges to $1 - g$ as skill rises without limit.

Assuming a constant elasticity of labor supply, e, and a Pareto distribution for skills above the mode with coefficient a, so that $C(n)$ equals $1/a$, (10') becomes

$$(14) \quad T'/(1-T') = (e^{-1}+1)B(n)/a.$$

Solving for T' and taking the limit as n rises, one has:

PROPOSITION 2: *Assuming a Pareto distribution of skills above the modal skill and a constant elasticity of labor supply, as skill rises without limit the optimal tax rate converges to*

$$(15) \quad T' = (e^{-1}+1)(1-g) \div [a + (e^{-1}+1)(1-g)].$$

[6] In this case, the tax rate tends to the revenue-maximizing rate, since, in the limit, the only effect of taxes on welfare is through the budget constraint.

TABLE 1—ASYMPTOTIC MARGINAL TAX RATES

	$g = 0$			$g = 0.25$			$g = 0.5$		
$a =$	0.5	1.5	5.0	0.5	1.5	5.0	0.5	1.5	5.0
e									
0.2	92	80	55	90	75	47	86	67	38
0.5	86	67	38	82	60	31	75	50	23

Notes: Asymptotic marginal tax rates, in percent, with a constant elasticity of labor supply, e, a Pareto distribution of skills with parameter a, and a ratio of social marginal utility with infinite income to average social marginal utility of g.

To examine the implications for the taxation of very high earners, values need to be selected for a, e, and g. Identifying skill with the wage yields an elasticity based on adjusting hours of labor supply; identifying skill with an underlying ability suggests a larger elasticity since education is also variable. With a zero income effect, compensated and ordinary labor supply elasticities are the same. Presumably it is the compensated elasticity that one would want to use for illustrative purposes. Recognizing that I am seeking an elasticity for high earners, looking at the elasticity for prime-age males provides an approximation. Based on the survey by John Pencavel (1986), calculations are done for a range of elasticities from 0.2 to 0.5. A range of g from 0 to 0.5 seems very wide. For the coefficient of the Pareto distribution, using tax data Daniel R. Feenberg and James M. Poterba (1993) find a varying between 0.5 and 1.5 over the years 1951–1990 for the incomes of the top 0.5 percent of the population. The calculations reported below suggest the possibility of a considerably higher value for the distribution of skills, perhaps as large as 5. Values of the asymptotic marginal tax rate [from (15)] are shown in Table 1. Thus I conclude that there is a case for high marginal tax rates in the quasi-linear Mirrlees model with plausible empirical parameters.

V. Decreasing Marginal Tax Rates

In considering decreasing marginal tax rates, I consider only the levels of skills above n_C, the level at which G' equals p. At skill levels above n_C, it would be desirable to transfer resources away from this skill level (if it could be done costlessly). It is now convenient to work with equation (9), which I rewrite

$$(9)\quad T'/(1-T') = [e^{-1}+1]\left[\int_{n}^{n_1}(p-G')\,dF\right] \div [pnf(n)].$$

As noted in Lemma B, at skill levels above n_C, the integral in (9) is decreasing with skill. Below the mode, the density is rising and so $1/[nf(n)]$ is falling with skill. Thus with a constant or rising elasticity of labor supply, the marginal tax rate is declining with skill. This argument also goes through above the mode where $nf(n)$ is rising with skill. This is summarized in:

PROPOSITION 3: *Above the critical skill level, n_C, marginal tax rates are decreasing where the elasticity of labor supply is constant and the distribution of skills has $nf(n)$ rising with skill.*

While one would expect $nf(n)$ to be increasing in n just above the modal skill, empirically, this seems unlikely at high skills, as is indicated in the data discussed below.

One can also use (9) to compare tax rates at two income levels above n_C, on either side of the modal skill and such that the density is equal at the two points. With G' less than p at the lower of the two skill levels being compared, the marginal tax rate would be higher

at the lower income level with a constant or rising elasticity of labor supply. This is summarized in:

PROPOSITION 4: *Above the critical skill level, n_C, marginal tax rates are higher at the lower of two skill levels that have the same density and the same the elasticity of labor supply at the chosen labor supplies.*

Combining results, one can see the pattern of tax rates when the density of skills is single-peaked (and such that the workers with the modal skill work and have G' less than p in equilibrium). With a constant elasticity of labor supply and the Pareto distribution of skills where the density is falling, the pattern of marginal tax rates is U-shaped above n_C, with the minimum of marginal rates occurring at the modal skill. Moreover, marginal rates are higher at the lower income levels. Plausibly, the density of skills does not have a kink at the mode, but changes smoothly from rising to declining as a Pareto density. Then, with the conditions in Proposition 3 the minimum of the tax rate (over the range above n_C) occurs above the modal skill. It is worth reiterating that the range with declining marginal rates need not begin at zero earned income.

VI. Another Example

The assumption of a constant elasticity of labor supply relates the optimal tax to a familiar concept in the analysis of deadweight burdens. By moving the term "n" from $C(n)$ to $A(n)$, one finds another example with similar conclusions. Consider the logarithmic case, $v(1-y) = \log(1-y)$. In this case, the elasticity of labor supply is equal to $(1-y)/y$. Thus one has:

LEMMA A′: *If $v(1-y) = \log(1-y)$, then $A(n) = n(1-T')$.*

If, above the modal skill level, the distribution is the exponential distribution, then $nC(n)$ is a constant.

LEMMA C′: *For $n < n_m$, $nC(n)$ is decreasing in n. For $n > n_m$, $nC(n)$ is constant if $F(n)$ is the exponential distribution above n_m.*

I can now put together Lemmas A′, B, and C′.

PROPOSITION 1′: *Marginal tax rates are increasing above the modal skill if, above the skill, the utility-of-leisure is logarithmic and the distribution of skills is exponential.*

For Proposition 1′, it is noted that $T'/(1-T')^2$ varies with n as $B(n)$ does. With $B(n)$ increasing, so too is T'. As above, from the arguments that led to Proposition 1′, one can see that the result carries over if, at the equilibrium labor supplies, the elasticity of labor supply falls with skill more than in the stated condition. Similarly, it is sufficient to have a distribution of skills such that $[1-F(n)]/f(n)$ rises. Moreover, the result of rising tax rates will hold for part of the skill distribution (above the mode) if the conditions are met for that part; one does not need conditions on the entire distribution.

For the case just analyzed, the asymptotic marginal rate is calculated. With a logarithmic utility-of-leisure function and an exponential distribution of skills (above the mode) with coefficient b, (10) becomes:

$$(16) \quad T'/(1-T')^2 = B(n)(1-F)/f$$

$$= B(n)/b.$$

Solving for T' and taking the limit, one has:

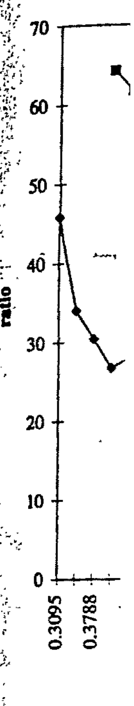
PROPOSITION 2′: *Assuming an exponential distribution of skills above the modal skill with parameter b and logarithmic utility-of-leisure, as skill rises without limit the optimal tax rate converges to*

$$(17) \quad T' = 1 - [(b'^2 + 4b')^{1/2} - b']/2,$$

where $b' = b/(1-g)$.

For g equal to 0 and $(1-F)/f$, that is, $1/b$, of 5, 10, and 15 (see Figures 1 and 2), the optimal marginal tax rate tends to 64, 73, and 77 percent. For g equal to 0.5 and the same values of $1/b$, the optimal marginal tax rate tends to 54, 64, and 70 percent.

Similarly, one can examine conditions for declining marginal tax rates with the lo



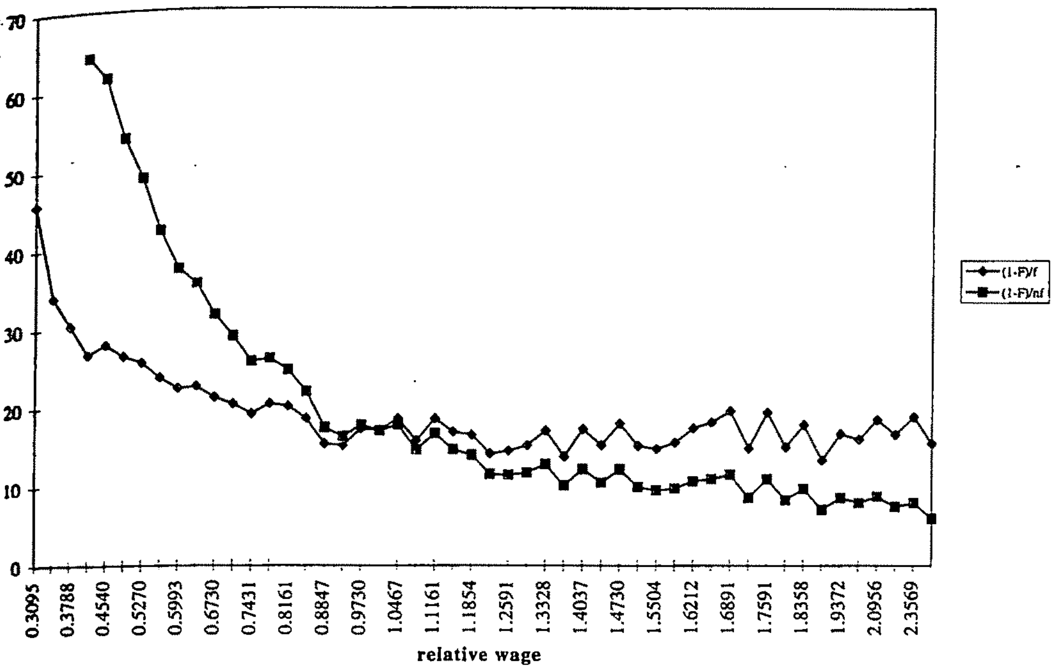


FIGURE 1. RATIOS $[1 - F(n)]/f(n)$ AND $[1 - F(n)]/[nf(n)]$ CALCULATED FROM RELATIVE WAGES

arithmic utility-of-leisure. In this case, (9) becomes:

$$(18)\quad T'/(1-T')^2 = \left[\int_n^{n_1} (p - G')\, dF\right] \Big/ [pf].$$

From Lemma B, it can be concluded that the tax rate is declining above n_C and below the mode.

PROPOSITION 3′: *Between the critical skill level, n_C, and the mode, n_m, marginal tax rates are decreasing if the utility-of-leisure is logarithmic.*

Similarly, from (18) one can conclude:

PROPOSITION 4′: *Above the critical skill level, n_c, marginal tax rates are higher at the lower of two skill levels that have the same density if the utility-of-leisure is logarithmic.*

VII. Data on the Distribution of Skills

While a careful attempt to fit this model to available data is beyond the scope of this paper, it does seem interesting to examine the distribution of wages. For this purpose, calculations have been done using the March 1992 CPS. This survey asked individuals for annual earnings in 1991, as well as weeks worked and typical hours per week. From these numbers one can calculate an implied average wage.[7] Using these wages, calculations were made of the mean wage per cell; the number of observations per cell, adjusted by interval width in order to be proportional to the density; and the number of observations with higher wages. Approximately 17 percent of the sample report wages below \$1 or no work and are omitted. In order to have reasonable cell sizes, the wage intervals are first \$0.50, but are expanded above a wage of \$26. As expected, a smoothing of the data would show a single-peaked distribution, as assumed in the analysis above. In Figure 1 is shown the ratios $(1 - F)/f$ and $(1 - F)/(nf)$, where n is measured as the wage relative to the mean wage. Because the series are very noisy, the graph is a centered three-cell moving average.

[7] No attempt was made to consider both earners in a two-earner family or wages of single females.

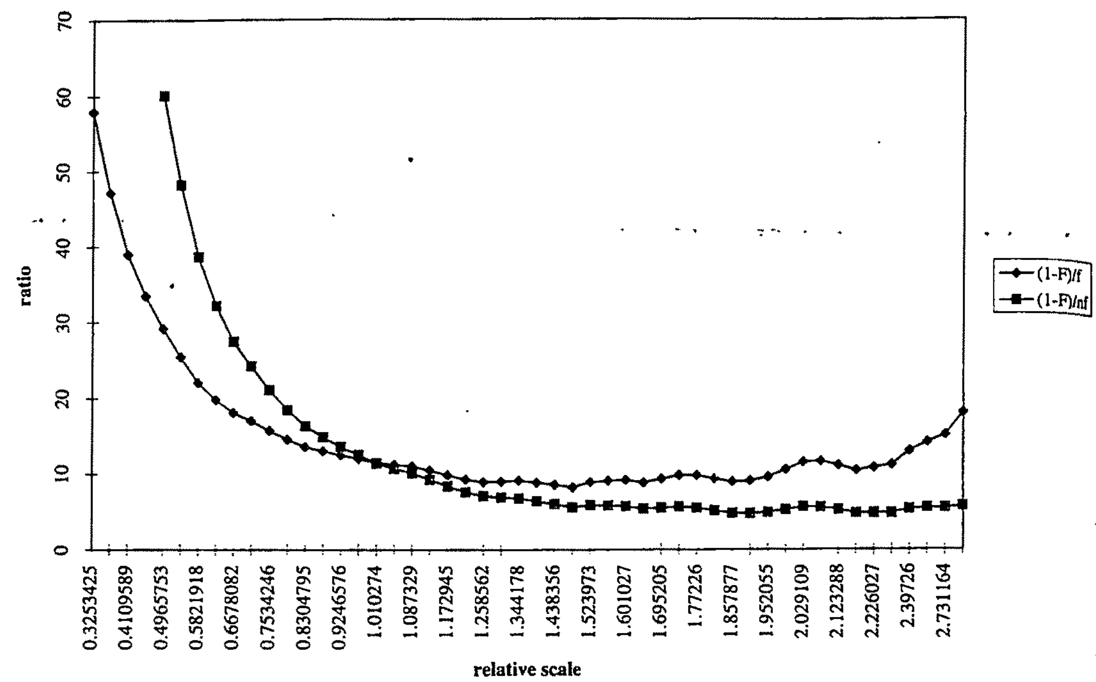


FIGURE 2. RATIOS $[1 - F(n)]/f(n)$ AND $[1 - F(n)]/[nf(n)]$ CALCULATED FROM RELATIVE SKILLS

For readability, the lowest wages are dropped from the graph since the ratios are very large. The figure shows sharply falling values of $(1 - F)/f$ through the range where the density is first rising and then roughly flat, that is, up to a wage of roughly \$13, a little below the mean of \$13.70. Beyond this point $(1 - F)/f$ is roughly constant at a value around 15. This implies a downward trend in $(1 - F)/(nf)$. A constant value of $(1 - F)/f$ is consistent with an exponential distribution over this range of values.

With a longer time horizon than one year, one would consider education to be an endogenous variable somewhat responsive to tax incentives. One might also be interested in the distribution of skills within a cohort. Thus a further calculation was done by regressing the log of the wage on education, age, and age squared, and plotting the exponentiated residuals. In Figure 2 are the same curves for this distribution as shown in Figure 1 for the distribution of wages (except that a moving average was not used). This distribution shows a fatter tail than the distribution of wages, with $(1 - F)/f$ rising and $(1 - F)/(nf)$ roughly constant for the top 15 percent of the skill distribution. A constant value of $(1 - F)/(nf)$ is consistent with a Pareto distribution over this range of values.

VIII. Concluding Remarks

The absence of income effects allows an intuitive grasp of the factors that determine the optimal tax structure. Increasing the marginal tax rate affecting some skill level involves an increase in the deadweight burden for people at this skill level. Thus, the optimal marginal tax rate at some income level depends on the elasticity of labor supply at that income level, since this is important for marginal distortions. Increasing the marginal tax rate also transfers income from all individuals with higher skills to the government, without changing the distortions of their labor supplies. The weights on these two elements depend on the ratio of individuals with skills above this level to individuals with skills at this level and on the level of skill which links the tax on hours to the tax on income. This intuition is displayed in the equations above, where the first-order condition for the optimal income tax is written as a product of these three terms.

The rewriting of the first-order condition also highlights the critical role of the assumed family of distributions for the upper tail of skills, as opposed to just the value of the parameters. With the Pareto distribution, $(1 - F)/(nf)$ is a constant and the change in the marginal tax rate reflects the rate of decline in social marginal utility of income as well as the change in labor supply elasticity with skill. With the exponential distribution, $(1 - F)/(nf)$ declines at the rate $1/n$. Thus either the elasticity of labor supply or the social marginal utility of income needs to be falling sufficiently rapidly to have constant or rising marginal tax rates. In the simulations in Mirrlees (1971), it was assumed that the distribution of skills was lognormal, so that $(1 - F)/(nf)$ declines at the rate $1/\log(n)$. Presumably the relatively constant marginal tax rate in the those simulations would have had a different shape with a different assumed family of distributions. Exploration of the shape of this distribution is clearly important for the normative case for different degrees of income tax progressivity.

There is not a simple route between the Mirrlees model and policy implications for annual income taxes levied repeatedly on families and covering both capital and labor incomes. The assumption of a zero income elasticity of labor supply and the limited information on both the shape of the skill distribution and the pattern of elasticities of labor supply by skill level would limit inferences even if there were a simple route. Nevertheless there are some lessons from the analysis. The sharp fall in $(1 - F)/f$ as skills approach the mode of the skill distribution from below seems highly relevant, especially if one wants to redistribute from people near the mode, rather than to them. This finding on the shape of an optimal (negative) income tax seems relevant in thinking about the phaseout of the earned income tax credit, and, possibly, welfare reform. That is, labor supply depends on the net return to earnings, which, in turn, depends on both the income tax and the phaseout of income-tested benefits. The presence of high marginal tax rates in the region of phaseout of benefits is not necessarily a basis for criticism of the programs—the optimal program may well have such a shape because of the advantage of higher marginal rates over a shorter range of skills where the skill density is large and rising. In other words, a sizable implicit marginal tax rate where benefits are being phased out is consistent with the U-shaped pattern of marginal rates and may well be optimal.

Second, this model confirms the implication of Mirrlees' calculations that the optimality of a zero tax rate at the highest income level is not a finding that sheds much light on optimal taxes, especially in the absence of knowledge of exactly where the top is. That is, if one replaced an unbounded distribution of skills by a bounded one with the same distribution up to some level and a concentration of skills at the highest levels, the result of rising marginal tax rates continues to hold until the concentration at the top is reached. There is no need for tax rates to decline slowly toward zero as one approaches the absolute top of the skill distribution.

Third, the sensitivity of the pattern of marginal rates to the measure of skill seems relevant, although different formulations of "skill" will be associated with different estimates of the elasticities of labor supply as well as different estimates of the shape of the distribution of skills. This analysis emphasizes the importance of the shape of the distribution of skills for optimal tax rates.

APPENDIX: HEURISTIC DERIVATION OF THE OPTIMAL TAX FIRST-ORDER CONDITION

Using just the first-order condition for labor supply for the quasi-linear utility function as a constraint, the optimal tax problem is

$$\text{(A1)} \quad \operatorname{Max} \int_{n_0}^{n_1} G\{u[x(n), y(n)]\} f(n)\, dn,$$

$$\text{subject to: } \int_{n_0}^{n_1} x(n) f(n)\, dn$$

$$\leq \int_{n_0}^{n_1} n y(n) f(n)\, dn - E;$$

$$u'[1 - y(n)] = n(1 - T'[ny(n)]).$$

With $x(n) = ny(n) - T[ny(n)]$, and using the first-order condition for labor supply, the change in consumption with skill satisfies:

$$\text{(A2)} \quad x'(n) = y(n)(1 - T') + n(1 - T')y'(n) = [y(n) + ny'(n)]v'/n.$$

With the quasi-linear utility function, one can calculate the derivative of u with respect to n:

$$\text{(A3)} \quad u'(n) = x'(n) - v'y'(n) = y(n)v'/n.$$

Treating $u(n)$ as a state variable and $y(n)$ as a control variable, the optimal tax problem can be rewritten as

$$\text{(A4)} \quad \text{Max} \int_{n_0}^{n_1} G[u(n)]f(n)\,dn,$$

subject to:

$$\int_{n_0}^{n_1} \{u(n) - v[1 - y(n)]\}f(n)\,dn \leqq \int_{n_0}^{n_1} ny(n)f(n)\,dn - E;$$

$$u'(n) = y(n)v'[1 - y(n)]/n.$$

Forming a Hamiltonian for this expression,

$$\text{(A5)} \quad H = \{G[u(n)] - p[u(n) - v[1 - y(n)] - ny(n)]\} \times f(n) + h(n)y(n) \times v'[1 - y(n)]/n,$$

where p and $h(n)$ are multipliers. The derivative of h is equal to minus the partial derivative of the Hamiltonian with respect to u:

$$\text{(A6)} \quad h'(n) = -\{G'[u(n)] - p\}f(n).$$

Maximizing the Hamiltonian with respect to $y(n)$,

$$\text{(A7)} \quad -p\{n - v'[1 - y(n)]\}f(n) = h(n)\{v'[1 - y(n)] - y(n)v''[1 - y(n)]\}/n.$$

Recognizing that $h(n_1)$ is equal to zero, (A6) can be integrated from n to n_1 to have an expression for $h(n)$. Substituting in (A7), one then has the first-order condition in the text (8).

REFERENCES

Atkinson, Anthony B. "Public Economics and the Economic Public." *European Economic Review*, May 1990, *34*(2–3), pp. 225–48.

Ebert, Udo. "A Reexamination of the Optimal Nonlinear Income Tax." *Journal of Public Economics*, October 1992, *49*(1), pp. 47–73.

Feenberg, Daniel R. and Poterba, James M. "Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns," in James Poterba, ed., *Tax policy and the economy*. Cambridge, MA: MIT Press, 1993, pp. 145–77.

Feldstein, Martin. "The Optimal Level of Social Security Benefits." *Quarterly Journal of Economics*, May 1985, *100*(2), pp. 303–20.

***Financial Times*.** "The Markets: Still Waiting for Answers on Tax—Economics Notebook." September 11, 1995, p. 24.

Fudenberg, Drew and Tirole, Jean. *Game theory*. Cambridge, MA: MIT Press, 1991.

Kanbur, Ravi and Tuomala, Matti. "Inherent Inequality and the Optimal Graduation of Marginal Tax Rates." *Scandinavian Journal of Economics*, June 1994, *96*(2), pp. 275–82.

Lollivier, Stefan and Rochet, Jean-Charles. "Bunching and Second-Order Conditions: A Note on Optimal Tax Theory." *Journal of Economic Theory*, December 1983, *31*(2), pp. 392–400.

Mirrlees, James A. "An Exploration in the Theory of Optimum Income Taxation." *Review of Economic Studies*, April 1971, *38*(114), pp. 175–208.

Myles, Gareth D. *Public economics*. Cambridge: Cambridge University Press, 1995.

Pencavel, John. "Labor Supply of Men: A Survey," in Orley Ashenfelter and Richard Layard, eds., *Handbook of labor economics*. Amsterdam: North-Holland, 1986, pp. 3–102.

Sadka, Efraim. "On Income Distribution, Incentive Effects and Optimal Income Taxation." *Review of Economic Studies*, June 1976, *43*(2), pp. 261–67.

Seade, Jesus K. "On the Shape of Optimal Tax Schedules." *Journal of Public Economics*, April 1977, *7*(2), pp. 203–35.

———. "On the Sign of the Optimum Marginal Income Tax." *Review of Economic Studies*, October 1982, *49*(4), pp. 637–43.

Tuomala, Matti. "On the Optimal Income Taxation: Some Further Numerical Results." *Journal of Public Economics*, April 1984, *23*(3), pp. 351–66.

———. *Optimal income tax and redistribution*. Oxford: Clarendon Press, 1990.

Weymark, John A. "Comparative Static Properties of Optimal Nonlinear Income Taxes." *Econometrica*, September 1987, *55*(5), 1165–85.

Endogenously Chosen Boards of Directors and Their Monitoring of the CEO

By BENJAMIN E. HERMALIN AND MICHAEL S. WEISBACH*

How can boards be chosen through a process partially controlled by the CEO, yet, in many instances, still be effective monitors of him? We offer an answer based on a model in which board effectiveness is a function of its independence. This, in turn, is a function of negotiations (implicit or explicit) between existing directors and the CEO over who will fill vacancies on the board. The CEO's bargaining power over the board-selection process comes from his perceived ability relative to potential successors. Many empirical findings about board structure and performance arise as equilibrium phenomena of this model. (*JEL* D23, D73, G39, K22, L29)

Corporations are not governed by the process that corporate law would seem to imply. Corporate law states that shareholders choose the board of directors, but, in practice, shareholders almost always vote for the slate proposed by management.[1] Moreover, this slate is approved by, if not chosen by, the very CEO these directors are supposed to monitor (see, e.g., Myles L. Mace, 1971; Jay W. Lorsch and Elizabeth MacIver, 1989; Ada Demb and F.-Friedrich Neubauer, 1992). The resulting governance system has been criticized as ineffective in controlling management (see, e.g., Martin Lipton and Jay W. Lorsch, 1992; Michael C. Jensen, 1993).

Given these apparent shortcomings, it is easy to forget that the current system is, nonetheless, the market solution to an organizational design problem (albeit one that must be solved under legal constraints—e.g., all firms must have boards with certain powers).[2] Thus, as George J. Stigler and Claire Friedland (1983) argued, before any criticism of current practice is taken too seriously, a thorough understanding of the market forces that have led to its existence seems necessary. This, in part, is what we propose to do here.

The previous literature has focussed on what boards do, without asking how they got to be the way they are. However, the answers to these questions are invariably linked. For instance, a board packed with the CEO's relatives will be less effective than one made up of large shareholders. To understand corporate governance, the questions of director choice and director function must be answered simultaneously.

To capture this simultaneity, we assume that the board and the CEO negotiate over both the

* Hermalin: Haas School of Business and Department of Economics, S545 Student Services Building No. 1900, University of California, Berkeley, CA 94720; Weisbach: Department of Finance, College of Business and Public Administration, McClelland Hall, Room 315R, University of Arizona, Tucson, AZ 85721. The authors thank James Brickley, Aaron Edlin, Robert Gibbons, Charles Hadlock, David Hirshleifer, Steven Kaplan, Vojislav Maksimovic, Canice Prendergast, Anil Shivdasani, Sunil Wahal, Jerry Weisbach, two anonymous referees, and seminar participants at the University of Alberta, University of Arizona, Boston College, University of California-Berkeley, UCLA, University of California-San Diego, University of Chicago, Cornell University, Dartmouth College, Hong Kong University of Science and Technology, University of Illinois, Indiana University, University of Maryland, University of Rochester, and the Winter Econometric Society for their helpful comments and suggestions on earlier drafts.

[1] Even when shareholders do challenge management's slate of directors in a proxy fight, Harry and Linda DeAngelo (1989) find that they win a board seat only about one-third of the time.

[2] Mark Roe (1994) expands on the role of legal and political constraints. These, however, are better suited to explaining cross-country differences in corporate governance than the intracountry differences that are our focus.

CEO's wage and the identity of new directors. These negotiations could be explicit. Alternatively, in keeping with the institutional literature (see, e.g., Mace; Lorsch and MacIver; Demb and Neubauer), these negotiations could be implicit—the CEO could nominate new board members subject to a tacit understanding about the set from which they may be chosen. Were the CEO to violate this understanding, the board would refuse to approve his nominees. The CEO's bargaining power in these negotiations comes from his *perceived* ability relative to a replacement.

These negotiations determine the board's level of independence. Independence is important because a director's willingness to monitor the CEO increases with his or her independence. Monitoring provides information about the CEO used by the board in deciding whether to retain or to replace him. In this model, therefore, both the structure of the board and its actions are endogenously derived.

To evaluate the model's realism, we compare its predictions to existing empirical findings. Some of the model's predictions are:

1. A CEO who performs poorly is more likely to be replaced than one who performs well.
2. CEO turnover is more sensitive to performance when the board is more independent.
3. The probability of independent directors being added to the board rises following poor firm performance.
4. Board independence declines over the course of a CEO's tenure.
5. Accounting measures of performance are better predictors of management turnover than stock-price performance.

These predictions are consistent with existing studies of large corporations.

Predictions 1 and 2 match the empirical evidence on CEO turnover. A number of papers (e.g., Anne T. Coughlan and Ronald M. Schmidt, 1985; Jerold B. Warner et al., 1988; Jensen and Kevin J. Murphy, 1990), find that CEO turnover is negatively related to prior performance. In addition, Weisbach (1988) finds that the sensitivity of CEO turnover to performance is greater for firms with a higher proportion of outside directors, presumably because these firms' boards are more independent of management than are boards dominated by inside directors.[3]

While prediction 1 follows from other models (e.g., David Hirshleifer and Anjan V. Thakor, 1994), prediction 2 is more novel. The intuition behind it is that new CEOs require more monitoring than old ones, since less is known about their ability. More independent boards have a greater tolerance for this added monitoring, so they can afford to be tougher with an incumbent CEO whose performance is marginal.

Prediction 3 is consistent with the pattern of director turnover found by Hermalin and Weisbach (1988). They also find that the proportion of outside directors on the board decreases over the CEO's career. This finding suggests that board independence declines over a CEO's tenure, consistent with prediction 4.

Prediction 3 follows because poor performance lowers the board's assessment of the CEO's ability, reducing his bargaining position and thus increasing the probability that the CEO will be forced to accept more independent directors. Similar logic explains prediction 4: If a CEO keeps his job, then retaining him must be worth more to the directors than replacing him. This means that this CEO is, to some extent, a rare commodity, which gives him bargaining power vis-à-vis the directors. He is, therefore, able to bargain for a board that is more favorable to him.

Prediction 5 is consistent with both Weisbach (1988) and Murphy and Jerold L. Zimmerman (1993), who estimate equations predicting management changes using both stock returns and earnings and find that earnings do a better job. Intuitively, earnings are a function of current management only, but stock returns reflect both current management and the expectation of future management changes.

[3] Outside directors are nonmanagement directors not otherwise affiliated with the firm and inside directors are management or directors with close ties to the firm (e.g., ex-CEOs of the firm). Some researchers also include a third category, "grey" or "affiliated," intermediate in presumed independence between outsiders and insiders.

In addition, the analysis suggests the following predictions, which have not yet been empirically tested.

1. There should be long-term persistence in corporate governance. In particular, changes that either strengthen or weaken board independence should be "permanent" in that they change the long-term bargaining strength of the board against management.
2. The stock-price reaction to management changes should be negative if the manager is fired on the basis of private information and positive if the manager is fired on the basis of public information.
3. A CEO's salary should be insensitive to past performance at relatively low levels of past performance, but sensitive at relatively high levels of past performance.

The first prediction is consistent with anecdotal evidence of long-term patterns in corporate governance. When a firm has an extremely able CEO, he will be able to use his bargaining position to ensure a relatively weak board throughout his career. Consequently, his successor will inherit a relatively weak board. Thus, the model suggests that there will be long-term persistence in firms' governance practices and long-term interfirm heterogeneity in these practices as well.

The second prediction potentially explains why empirical studies have found no consistent stock-price reaction to management changes (see Warner et al.): there have been no controls for whether the dismissals were due to private or public information. Our prediction follows because a change in management conveys information about both the board and the CEO. If the board bases its firing decision on private information, then a firing reveals that a CEO who was previously seen as better than the expected value of a replacement is not. Expectations of new management are lower than the previous expectations of old management, so the stock price falls. In contrast, if the firing is based on public information, then nothing new is revealed about the CEO, but the firing conveys good news about the board's independence, so the stock price rises.

The third prediction comes from the structure of the bargaining game in our model. When the CEO is either new or a mediocre performer, the lower bound on his wage binds (a bound stemming from a limited-liability assumption). As performance increases, the CEO's bargaining position increases as well, allowing him to capture a fraction of the rent in the form of a higher wage.

Although we focus on explaining phenomena related to boards of directors, the model we develop is fairly general. It extends the job-matching model of Boyan Jovanovic (1979) by allowing for *endogenous* monitoring decisions. Among its features is a formalization of the argument that new workers are more valuable than older, better-known workers, ceteris paribus, because the former have a greater option value (see Proposition 2 below).

The paper has the following organization. The next section reviews some previous work on boards of directors. Section II introduces our model. We extend it in Section III, subsection A, by allowing the board to prefer the incumbent CEO to replacements. In Section III, subsection B, we consider how various governance activities affect the firm's stock price and how measures of firm performance will correlate with governance activities. Section IV considers a reinterpretation of our model that eschews bargaining and ensures that turnover is always optimal from the shareholders' perspective. Many of our results continue to hold under this alternative interpretation, although we lose the ability to analyze management capture of the board. Section V considers some policy prescriptions that have been offered to correct the perceived failings of boards (e.g., requirements that directors be paid in stock rather than cash). Our model predicts that many of these policies will be ineffective. We conclude in Section VI.

I. Boards of Directors in Corporate Governance: Existing Theory and Evidence

Adolph A. Berle, Jr. and Gardiner C. Means (1932 p. 87) observed that the separation of ownership and control inherent in a diversely held corporation leads to a board of directors controlled by management. They argued that

> ... control will tend to be in the hands of those who select the proxy committee

> and by whom, the election of directors for the ensuing period will be made. Since the proxy committee is appointed by the existing management, the latter can virtually dictate their own successors.

Management's apparent control of the board selection process led Berle and Means, as well as more contemporary authors such as Jensen, to question whether directors can be effective monitors.

A counterargument is that directors' concern for their reputations in the managerial or directorial labor market causes them to be effective monitors (Eugene F. Fama, 1980; Fama and Jensen, 1983). Steven N. Kaplan and David Reishus (1990) find evidence consistent with this argument: directors of poorly performing firms, who therefore may be perceived to have done a poor job overseeing management, are less likely to become directors at other firms. On the other hand, as Bengt Holmstrom (1983) shows, reputational concerns need not correct all agency problems and can, in fact, create new ones.

To resolve these conflicting arguments about board effectiveness, an empirical literature assessing the board's role has developed. Unfortunately, measuring the day-to-day effect of board independence on corporate profits is made difficult by simultaneity problems, since board independence itself changes because of corporate performance (see Hermalin and Weisbach, 1991, for further discussion).[4] This literature has been more successful analyzing the relation between corporate decision-making in specific circumstances and some measure of the board's independence, such as its composition in terms of insiders and outsiders or other characteristics correlated with independence (e.g., directors' shareholdings). For instance, Brickley et al. (1994a) find that board composition affects their decision to adopt poison pills; John W. Byrd and Kent A. Hickman (1992) find that the greater the fraction of outside directors, the better the stock market's reaction to their firm's tender offers for other firms; and Weisbach finds that the sensitivity of CEO turnover to firm performance increases with the fraction of outside directors on the board. Anil Shivdasani (1993) suggests that outside directors who own a substantial number of shares and who hold more corporate directorships (presumably measuring the value they place on their reputations) are better at negotiating a favorable deal for shareholders who face a takeover bid. Lastly, Kevin Hallock (1997) finds that firms whose boards are interlocked (contain a CEO on whose board the firm's CEO serves) tend to pay their CEOs more. He argues that interlocked directors are less independent and, consequently, give the CEO a larger fraction of the rents than would other directors.

While the empirical literature on boards is fairly well developed at this point, there has been little formal modeling of the board. Existing papers have considered important issues about the board's conduct, but have ignored the process by which boards get to be the way they are (see Hirshleifer and Thakor, 1994; Vincent A. Warther, 1994; Brickley et al., 1994b; Thomas H. Noe and Michael Rebello, 1996). Our view is that such modeling is useful, but it ignores the fundamental question of how boards can, in many instances, be effective monitors of the CEO despite being chosen through a process partially controlled by him.

II. The Model

We model the board selection process as a bargaining game between the CEO and the board. We assume no active role for the shareholders (although some of them could be directors). Consistent with practice, shareholders simply ratify the slate put forward by the company. We discuss shareholder activity, particularly in crisis situations, later.

[4] Some papers that nonetheless attempt to estimate the day-to-day effect are Paul MacAvoy et al. (1983) and Barry D. Baysinger and Henry N. Butler (1985). Also see James A. Brickley and Christopher M. James (1987), Stuart Rosenstein and Jeffrey G. Wyatt (1990), William Brown and Michael T. Maloney (1992), Omesh Kini et al. (1994), Anup Agrawal and Charles R. Knoeber (1996), Sanjai Bhagat and Bernard Black (1996), James Booth and Daniel N. Deli (1996), Kenneth A. Borokhovich et al. (1996), David Yermack (1996), James F. Cotter et al. (1997), Robert Gertner and Kaplan (1997), David Mayers et al. (1997), and April Klein (1998) for related empirical work.

A. *Timing*

The game has multiple stages with the following timing.

1. At the start of the game, the firm has a new CEO. The *commonly* held prior distribution about his ability, α, is normal with mean zero and variance $1/\tau_0$ (τ_0 is the *precision* of the distribution). We set the mean to zero for convenience, but without loss of generality.
2. The first realization of earnings, x_1, occurs. Earnings are distributed normally with a mean equal to the CEO's true ability, α, and a variance equal to $1/r$.
3. Based on x_1, the board updates its estimate of the CEO's ability. The board may at this stage decide to fire the CEO and hire a replacement. The prior distribution of *any* replacement CEO's ability is normal with mean zero and variance $1/\tau_0$.
4. The CEO (either the incumbent or the replacement) negotiates with the board over filling vacancies on the board and his wage, w.[5] If the bargaining is unsuccessful, the CEO is fired and a replacement is hired. If a replacement is hired at this stage, the board bargains with him about the filling of vacancies on the board.
5. The board may then acquire a private signal, y, about the CEO. The probability that the board acquires this signal depends on the intensity with which it monitors the CEO. The signal is distributed normally with a mean equal to the CEO's ability, α, and a variance equal to $1/s$.[6]
6. If the board acquires the signal, it updates its estimate of the CEO's ability. Based on this posterior estimate, the board may decide to fire the CEO and hire a replacement.
7. The second realization of earnings (profits gross of the CEO's compensation), x_2, occurs. Again, earnings are distributed normally with a mean equal to the CEO's true ability and a variance equal to $1/r$. The random variables $y - \alpha$, $x_1 - \alpha$, and $x_2 - \alpha$ are independently distributed.

B. *Preferences and Ability*

The CEO in charge at stage 7 receives a control benefit of $b > 0$. A CEO who is dismissed prior to this stage (or not hired) receives a benefit of 0.

The CEO is also compensated with a wage, w, determined by the bargaining process between him and the board. This wage is paid *regardless* of whether the CEO survives to stage 7. A critical assumption is that the CEO is protected by limited liability; specifically, the wage must be nonnegative (i.e., $w \geq 0$).

A CEO's ability is fixed throughout his career. We follow Holmstrom (1983) by assuming that the CEO, like the board, knows only the *distribution* of his ability (i.e., that it is a normal distribution with mean zero and precision τ_0). We justify this assumption by noting that the uncertainty about a CEO's ability in a *particular* job is largely uncertainty about the match between him and the firm, which is similarly unknown to both the board and the CEO.

We assume that each director, i's, utility is

$$(1) \qquad \theta_i x_2 - \kappa_i d(p).$$

The constant $\theta_i > 0$ equals the director's marginal utility from firm profits, x_2.[7] We imagine that directors put different weights on profits for two reasons. First, directors' incomes depend on their own shareholdings. Second, d

[5] We treat the creation of vacancies on the board as exogenous. This is fairly realistic, since many vacancies arise for presumably exogenous reasons such as death, illness, or reaching the customary retirement age. We also do not explicitly consider how the number of vacancies might limit the ability of the board to adjust board composition. This is with little loss of generality, since (i) only the proof of Proposition 6 depends on achieving an interior solution (i.e., one in which vacancy limits do not bind); and (ii) board size often changes (see Hermalin and Weisbach, 1988, for evidence).

[6] Alternatively, we could assume that the board always receives a signal, but its precision (i.e., s) is endogenous (we thank Canice Prendergast for this point). This alternative leads to similar results, except the comparative statics with respect to τ are ambiguous.

[7] First-stage profits, x_1, are sunk by the time the directors act, so we need not explicitly account for their impact on directors' utilities.

directors' concerns for building reputations as competent managers may vary.

The variable p is the probability that the board obtains an additional signal, y, about the CEO. It reflects the intensity with which the board monitors the CEO. The disutility of monitoring is $\kappa_i d(p)$, where $d(\cdot)$ is a common, strictly increasing, strictly convex, and twice-differentiable function and κ_i is the director's distaste for monitoring. We imagine that directors' distastes for monitoring vary for three reasons. First, inside directors' careers are tied to the CEO's, so they rarely find it in their interest to monitor him. Second, the opportunity cost of the directors' time will vary among outside directors. Finally, directors who value the opportunity to serve on other boards could have an incentive to establish reputations for not "rocking the boat"; i.e., for not intensely monitoring the CEO.

Since utility functions are defined up to an affine transformation only, we can replace (1) with

$$x_2 - k_i d(p), \tag{2}$$

where $k_i = \kappa_i/\theta_i$. We interpret k_i as a measure of director i's *lack* of independence, at least in terms of the way he or she behaves.

We assume k_i is *fixed* for a given director i (at a specific firm and for a specific set of pecuniary incentives). In particular, it is invariant with respect to who the CEO is. We relax this assumption in Section III, subsection A, where we show it does not change the qualitative nature of our results. In addition, we discuss below the implications of basing individual director monitoring costs on the group level of monitoring.

Updating Beliefs and Optimal Monitoring

When new information is observed, either profits or a signal, the players update their beliefs about the CEO's ability. Specifically, if $\hat{\alpha}$ and τ are the prior estimates of the mean and precision of the distribution of the CEO's ability, then the posterior estimates are

$$\hat{\alpha}' = \frac{\tau\hat{\alpha} + tz}{\tau + t} \text{ and } \tau' = \tau + t, \tag{3}$$

where z is either x_1 or y and t is either r or s (see Morris H. DeGroot, 1970 p. 167). The posterior distribution is also normal.

From (3), the board has a more precise estimate of an *incumbent* CEO's ability at stage 3 than it would of any *replacement* CEO it hires. That is,

$$\tau > \tau_0, \tag{4}$$

where τ is the precision of its estimate of the incumbent CEO's ability. Intuitively, an incumbent is a "known entity," so there is less uncertainty about him than there would be about a new CEO.

The distribution of the signal y given the CEO's *true* ability, α, is normal with mean α and variance $1/s$; hence, the distribution of y given the CEO's *estimated* ability, $\hat{\alpha}$, is normal with mean $\hat{\alpha}$ and variance $1/s + 1/\tau$.[8] Define

$$H = \frac{s\tau}{s + \tau}$$

to be the precision of y given $\hat{\alpha}$.[9]

Observe that the board's posterior estimate of a CEO's ability is also the expected value of x_2. After fixing (sinking) the CEO's wage, it is also the expected value of profits.

The alternative to retaining a CEO is to hire a replacement. The expected earnings from a replacement are, by assumption, zero. Moreover, because all replacements are *ex ante* identical, they have no bargaining power. Hence, the directors can set a minimum wage, $w = 0$. The expected profit from a *replacement* CEO is, therefore, zero. Subsequent to obtaining a signal, y, the incumbent CEO will thus

[8] The random variable $y - \hat{\alpha}$ is the sum of two independently distributed normal variables: $y - \alpha$ and $\alpha - \hat{\alpha}$; hence $y - \hat{\alpha}$ is also normally distributed. Since the means of these two random variables are both zero, the mean of y given $\hat{\alpha}$ is, therefore, $\hat{\alpha}$. The variance of the two variables are $1/s$ and $1/\tau$ respectively, so the variance of $y - \hat{\alpha}$ and, thus, y given $\hat{\alpha}$ is $1/s + 1/\tau$.

[9] As a convention, we will denote functions of many variables, such as H, by capital letters. When we need to be explicit about an argument of such functions—for example, the function F evaluated at $x = x'$—we will write $F_{x=x'}$.

be dismissed if $\hat{\alpha}' < 0$. Using (3), we can restate the dismissal condition as

$$(5)\qquad -\frac{\tau\hat{\alpha}}{s} > y.$$

The firm's expected value *if* it will learn y is

$$V \equiv \int_{-\infty}^{\infty} \max\left\{0, \frac{\tau\hat{\alpha} + sy}{\tau + s}\right\}\sqrt{\frac{H}{2\pi}}\, e^{-(H/2)(y-\hat{\alpha})^2}\, dy.$$

Since the option to fire the CEO is a valuable option, it follows that $V > \hat{\alpha}$ for all τ.

Straightforward calculations reveal that V can be written as

$$(6)\qquad V = \hat{\alpha}\Phi(-(Y_c - \hat{\alpha})\sqrt{H}) + \frac{\sqrt{H}}{\tau}\phi((Y_c - \hat{\alpha})\sqrt{H}),$$

where Y_c is the left-hand side of (5), $\Phi(\cdot)$ is the distribution function of a standard normal random variable (i.e., with mean zero and variance one), and $\phi(\cdot)$ is its corresponding density function. Note that

$$\Phi(-(Y_c - \hat{\alpha})\sqrt{H})$$

is also the probability that the CEO will be retained if evaluated.

A higher-ability CEO is always better, but the value of the option to fire him is decreasing in $\hat{\alpha}$:[10]

LEMMA 1: *V is increasing in $\hat{\alpha}$, while $V - \hat{\alpha}$ is decreasing in $\hat{\alpha}$.*

That is, the value of additional information about the CEO's ability is smaller the greater is the prior estimate of his ability.

Consider, now, the issue of how the board decides on the intensity (probability, p) with which to monitor the CEO. We assume the board chooses p to maximize:

$$(7)\qquad \max_{p \in [0,1]} \; pV + (1-p)\max\{0, \hat{\alpha}\} - \bar{k}d(p)$$

where $\bar{k}$ reflects, in some way, the collective preferences of the board (i.e., $\partial\bar{k}/\partial k_i \geq 0$ for all i and strictly positive for at least one i). For instance, $\bar{k}$ could be the average of the k_i's. More consistent, perhaps, with theories of voting, $\bar{k}$ could be the median k_i. Note that the resulting p will be Pareto optimal from the perspective of the board members.

The first-order condition for (7) is

$$(8)\qquad V - \max\{0, \hat{\alpha}\} - \bar{k}d'(p) = 0.$$

Expression (7) is concave in p, so (8) is sufficient as well as necessary. Define P^* to be the solution to (8). To keep the analysis straightforward, we consider only interior solutions (i.e., $P^* \in (0, 1)$). Corner solutions are a relatively simple extension. Properties of P^* are:

PROPOSITION 1: *The intensity with which the board monitors the CEO, P^*, is*

(i) decreasing with its prior estimate of the ability, $\hat{\alpha}$, if $\hat{\alpha} \geq 0$;
(ii) decreasing with the precision of its prior estimate, τ;
(iii) decreasing with its collective lack of independence, $\bar{k}$; but
(iv) increasing with the precision of the signal (i.e., s).

Intuitively, the more costly monitoring is to the board's members (or the less weight they place on the firm's profits), the greater is the marginal cost of monitoring, so they engage in less of it. The more able the board believes the CEO to be, the less valuable is the option to fire the CEO, so the board monitors less. The option to fire the CEO is similarly less valuable the less uncertainty there is in its prior estimate, so the board monitors less intensely when the CEO's ability is known more precisely. However, the option to fire the CEO is more valuable the greater is the precision of the signal, so the board

[10] All proofs are in the Appendix.

monitors more intensely when the signal is more informative.

Proposition 1 is consistent with the general perception that less-independent boards do less monitoring and that long-established CEOs (i.e., CEOs with high values of τ) receive less scrutiny. Being monitored increases the likelihood of being dismissed, so Proposition 1 is also consistent with the evidence in Weisbach, which suggests that outsider-dominated boards (which are presumably more independent) are more likely to fire a poorly performing CEO than insider-dominated (less-independent) boards.

D. *Negotiations Between the CEO and the Board*

When they enter into negotiations, the board brings

$$pV + (1-p)\max\{0, \hat{\alpha}\} - \bar{k}d(p) + R$$

in surplus to the bargaining table, where R is the share of b the board can expect to capture from a replacement CEO. However, given the limited-liability assumption ($w \geq 0$), the board cannot capture any share of b from a replacement CEO. Consequently, $R = 0$. The incumbent CEO brings his *expected* benefit,

$$(p\Phi(-(Y_c - \hat{\alpha})\sqrt{H}) + 1 - p)b,$$

to the table. So their joint surplus is

$$(9)\quad pV + (1-p)\max\{0, \hat{\alpha}\} - \bar{k}d(p) + (p\Phi(-(Y_c - \hat{\alpha})\sqrt{H}) + 1 - p)b.$$

Maximizing (9) with respect to p yields the first-order condition

$$(10)\quad V - \max\{0, \hat{\alpha}\} - \bar{k}d'(p) - (1-\Phi)b = 0.$$

Comparing (10) to (8), we see that the marginal benefit of monitoring is lower in (10) by $(1-\Phi)b$, which means that the level of monitoring that maximizes joint surplus (9) is lower than the level of monitoring that maximizes the board's expected utility (7). That is, if P^{**} is the solution to (10), then $P^{**} < P^{*}$.

It is worth considering why $P^{**} < P^{*}$. Part of the surplus that can be shared by the *incumbent* CEO and the board is the incumbent's chance of getting the control benefit, b. If he is fired, then this chance is lost; it goes to the *replacement* CEO. Moreover, limited liability prevents the board from recapturing it by setting a negative wage. Consequently, the marginal *joint* benefit of monitoring is reduced.[11]

We assume that the board and the CEO cannot contract *directly* on the probability that the board will evaluate the CEO (i.e., p). This assumption is consistent with the general perception that it is difficult for outside parties to verify how diligent the board is in its monitoring function (if it were easy for outside parties to verify the board's diligence, presumably the board would contract with the shareholders on this issue). This, however, creates a problem, because the board's private incentive is to choose a level of monitoring greater than that which maximizes *joint* surplus; i.e., the board would choose P^* instead of P^{**}.

The only way for the CEO and board to avoid this problem of too much monitoring is to lower the board's incentives to monitor by raising $\bar{k}$. We interpret the negotiations over $\bar{k}$ as decisions over factors likely to affect the independence of the board, such as board composition (e.g., proportion of insiders versus outsiders), board compensation, and so forth.

We assume that when the board negotiates with the CEO over $\bar{k}$ and w it cares about its utility only; that is, it does *not* consider the new (future) directors' utility in their negotiation.

Let $\bar{k}_0$ denote the collective lack of independence of the *continuing* directors.[12] If

[11] In many ways, the situation is similar to Philippe Aghion and Patrick Bolton's (1987) exclusive-dealing model. There, a retailer (our board) and a monopoly producer (our incumbent CEO) enter into an exclusive-dealing contract because of their concern that an entrant (our replacement CEO) will capture future surplus.

[12] Note the flexibility to change board composition comes from filling *exogenous* vacancies or *adding* directors to the board—no continuing director need leave to realize a change in board composition.

bargaining with the CEO yields a new board with a different lack-of-independence parameter, $\bar{k}_1$, then the *continuing* directors' expected utility is

$$(11) \quad P^*_{\bar{k}=\bar{k}_1}V + (1 - P^*_{\bar{k}=\bar{k}_1})\max\{0, \hat{\alpha}\} - \bar{k}_0 d(P^*_{\bar{k}=\bar{k}_1}) - \bar{w}.$$

Observe that the equilibrium probability of obtaining a signal, P^*, is a function of the *new* board's lack of independence, $\bar{k}_1$, not the continuing directors' lack of independence, $\bar{k}_0$.

We model the negotiations between the CEO and the board as a Nash bargaining game: the CEO and board agree to the level of independence (i.e., $\bar{k}$) and wage that maximize the product of their surpluses from trade. Provided the limited-liability constraint does not bind, the resulting composition (i.e., $\bar{k}$) will also maximize their joint surplus.[13] Assuming that the CEO has bargaining power is consistent with the institutional literature on boards, which suggests that CEOs, both in the United States and abroad, have considerable say over who is nominated for board positions. It is also consistent with the view that a CEO who has proven himself to be more valuable (in expectation) than any potential replacement should have some degree of bargaining power.

The surplus of the players is the difference between what they expect to receive if an agreement is reached and what they expect to receive if no agreement is reached. If no agreement is reached, the CEO leaves the firm—in which case his utility is 0. The CEO's surplus is, therefore,

$$P^*\Phi(-(Y_c - \hat{\alpha})\sqrt{H})b + (1 - P^*)b + w.$$

If no agreement is reached, the board hires a replacement CEO. Let U_0 be the board's expected utility if it hires a replacement (we will derive its value shortly—see Lemma 2 below). The board's surplus is, thus,[14]

$$P^*V + (1 - P^*)\hat{\alpha} - \bar{k}_0 d(P^*) - w - U_0.$$

Under Nash bargaining, the board and the CEO choose $\bar{k}$ and w to maximize

$$(12) \quad (P^*V(\hat{\alpha}, \tau) + (1 - P^*)\hat{\alpha} - \bar{k}_0 d(P^*) - w - U_0) \times (P^*\Phi(-(Y_c - \hat{\alpha})\sqrt{H})b + (1 - P^*)b + w).$$

To maximize (12), we need to know the value of U_0.

LEMMA 2: $U_0 = P_0 V_0 - \bar{k}_0 d(P_0)$, *where* V_0 *is V evaluated for a new CEO—i.e.,*

$$V_0 = \frac{\sqrt{H}}{\tau_0}\phi(0)$$

—and P_0 *solves the equation*

$$V_0 - \bar{k}_0 d'(p) = 0;$$

that is, P_0 *is the existing board's utility maximizing level of monitoring of a new CEO. Moreover, the wage paid a replacement CEO is zero.*

Intuitively, new CEOs have no bargaining power, since they all have equal expected value. Consequently, the board can set a minimum wage and get its most preferred level of independence, which is to replicate its current level (i.e., $\bar{k}_0$).

Recall our assumption that the board can choose to fire the CEO *prior* to bargaining. It might, at first, seem that the board would fire the incumbent CEO if and only if his estimated ability were less than the estimated ability of

[13] Other bargaining games would yield qualitatively similar results provided the CEO's bargaining power increases with his perceived ability.

[14] The reader will note that we have replaced max{0, $\hat{\alpha}$} with $\hat{\alpha}$. As we will show in Proposition 2, any incumbent CEO who is not fired prior to bargaining must have an estimated ability greater than zero.

a replacement (i.e., if and only if $\hat{\alpha} < 0$). This is not, however, true:

PROPOSITION 2: *A unique finite cutoff, $A_c > 0$, exists such that an incumbent CEO is fired prior to bargaining if and only if his estimated ability is less than A_c.*

Proposition 2 follows because the right to evaluate a CEO creates a valuable option. The value of this option increases with the prior uncertainty about the CEO (i.e., $1/\tau$). Consequently, its value is greater for a new CEO than for an incumbent CEO. A new CEO is, therefore, more desirable than an incumbent CEO ceteris paribus. So, an incumbent's estimated ability must be strictly greater than a new CEO's if he is to retain his job.

A natural question to ask is whether less-independent boards tolerate worse performance than do more-independent boards; that is, do greater values of $\bar{k}_0$ lead to lower values of A_c. The answer is yes:

PROPOSITION 3: *The minimum estimated ability for the incumbent CEO at which he will be retained, A_c, falls as the board becomes less independent; that is,*

$$\frac{\partial A_c}{\partial \bar{k}_0} < 0.$$

Remark 1: Since $\hat{\alpha}$ is a decreasing function of performance [recall (3)], Proposition 3 implies that CEO dismissals are more sensitive to (negative) firm performance when the board is more independent. As such, Proposition 3 is consistent with the evidence in Weisbach.

Proposition 3 holds because a replacement CEO must be monitored more than an incumbent CEO. The less independent the board (i.e., the greater the board's distaste for monitoring, $\bar{k}_0$), the greater is the cost to such a board from hiring a new CEO because he requires more monitoring. Hence, such a board is more willing to tolerate a mediocre CEO than is a more-independent board (i.e., one

Consider, now, bargaining between the board and an incumbent CEO who will be retained (i.e., one for whom $\hat{\alpha} > A_c$). Maximizing (12) with respect to $\bar{k}$ and w yields first-order conditions that are equivalent to

$$(13) \quad (V - \hat{\alpha} - \bar{k}_0 d'(p)) \times (p\Phi b + (1-p)b + w) + (p(V - \hat{\alpha}) + \hat{\alpha} - \bar{k}_0 d(p) - w - U_0) \times (\Phi - 1)b = 0 \text{ and}$$

$$(14) \quad p(V-\hat{\alpha}) + \hat{\alpha} - \bar{k}_0 d(p) - w - U_0 - (p\Phi b + (1-p)b + w) \leq 0.$$

Since $\hat{\alpha} > A_c$, the board's expected utility exceeds U_0, so the second line of (13) is negative. The first line must, therefore, be positive, which implies

$$(15) \qquad V - \hat{\alpha} - \bar{k}_0 d'(p) > 0.$$

From the first-order condition for the board's optimal p (8), condition (15) implies that $p < P^*_{\bar{k}=\bar{k}_0}$, which, from Proposition 1, implies $\bar{k} > \bar{k}_0$. We have, thus, established:

PROPOSITION 4: *If the continuing directors choose to retain the CEO, then the new board will have less independence than did the continuing directors (i.e., $\bar{k} > \bar{k}_0$).*

We emphasize the word "continuing" because the new board is less independent only relative to those directors who continue to serve. The proposition does not compare the new board with the previous board (i.e., the continuing *and* departed directors).[15] If, on the

[15] To illustrate this point, consider a hypothetical board with ten directors: five outsiders and five insiders. Suppose that $k_i = k_{out}$ for the outsiders and $k_i = k_{in}$ for the insiders, $k_{out} < k_{in}$. Suppose, too, that $\bar{k}$ equals the board's average k. This board's average k is $\frac{1}{2}k_{in} + \frac{1}{2}k_{out}$. Suppose two

other hand, "normal" attrition from the board leads to an average level of independence among the continuing directors that approximates the level of independence of the original board, then Proposition 4 suggests an explanation for the finding that boards become less independent over the career of the CEO (see Hermalin and Weisbach, 1988).

Proposition 4 suggests that corporate governance is subject to a stochastic form of "entropy": in the long run, boards will be relatively ineffective, consistent with the common complaint leveled against them. Proposition 4 is, however, subject to two caveats that potentially affect its interpretation. First, a key assumption behind this result is that the monitoring burden is shared equally by the directors (i.e., each must expend p). If the monitoring burden could be shared unequally—if for instance monitoring was a team production problem such as considered by Holmstrom (1982)—then this entropy result need not hold. Directors would have an incentive to free ride on the diligence of other directors. This in turn would give them an incentive to want new directors with strong proclivities for monitoring (i.e., low k_i's) rather than, as here, the same proclivity they have. In a richer model of board activity, entropy would, then, depend on the degree to which monitoring is a collective activity (as here) versus a private activity (as in a teams problem).

Second, we have assumed away any role for the shareholders, in keeping with the institutional evidence that they rarely play a direct role in either the "normal" selection of directors or the day-to-day operations of the company (see Mace, for example). Proposition 4 serves to emphasize the importance of those occasions when shareholders "break" the entropy through hostile takeovers, proxy fights, or direct negotiations between large shareholders and management.

Whether the bargaining maximizes the board and incumbent CEO's joint surplus depends on whether the limited-liability constraint binds. If it does not bind, then (14) is an equality. Using it, (13) becomes

$$V - \hat{\alpha} - \bar{k}_0 d'(p) - (1 - \Phi)b = 0,$$

which is the first-order condition for maximizing joint surplus (9). If the limited-liability constraint does bind, then (13) is equivalent to

$$V - \hat{\alpha} - \bar{k}_0 d'(p) - (1 - \Phi)\zeta b = 0,$$

where $\zeta < 1$. Consequently, the solution to the problem in which the limited-liability constraint is binding involves more monitoring and, hence, greater board independence than if the constraint is not binding. This establishes:

PROPOSITION 5: *Suppose that the incumbent CEO is retained. If the limited-liability constraint is not binding, then the level of monitoring will maximize the CEO and board's joint surplus. If it is binding, then the level of monitoring will exceed the joint-surplus-maximizing level. Correspondingly board independence will be greater if the constraint is binding than if it is not binding.*

We also want to know how estimated ability affects the ultimate equilibrium level of scrutiny (the probability of being evaluated) that the CEO will face.

PROPOSITION 6: *The equilibrium probability that the future board evaluates an incumbent CEO who is retained is decreasing with the prior estimate of his ability.*

Remark 2: Given the monotonic relationship between monitoring and board independence and between first-period earnings performance and estimated ability, Proposition 6 implies that performance and the independence of additions to the board should be negatively correlated, which is consistent with Hermalin and Weisbach's (1988) findings.

Propositions 4 and 6 show that history matters in corporate governance; that is, we c[illegible]

directors is $^3/_8 k_{in} + {^5/_8} k_{out}$. Suppose, consistent with Proposition 4, bargaining results in the addition of one insider and one outsider. The new board's average k is $^2/_5 k_{in} + {^3/_5} k_{out}$. The new board is, therefore, less independent than the continuing directors, but is more independent than the original board.

expect some hysteresis. Strong, independent boards will beget stronger, more-independent boards than will weak boards. Consequently, if we follow two firms, A and B, over time, then there is a good probability that we will find A always has a stronger board than B if A initially has the stronger board. Moreover, this could still be true even if B's recent performance is better than A's; indeed, from Proposition 6, better performance by B (i.e., a higher estimate of $\hat{\alpha}$) could actually accentuate the difference between the two boards' relative strength. These results underscore the importance of considering endogeneity in empirical studies of corporate governance and they potentially explain the inconclusive results of MacAvoy et al. and others.

Finally, we consider the relationship between the wage, w, and estimated ability.

PROPOSITION 7: *There exists an $\tilde{A}$, $A_c < \tilde{A}$, such that an incumbent CEO with estimated ability in $[A_c, \tilde{A}]$ is paid a wage of zero. There also exists an $\hat{A}$, $\tilde{A} < \hat{A} < \infty$, such that an incumbent CEO's wage is increasing in his estimated ability for estimated abilities in $[\hat{A}, \infty)$.*

For a retained CEO, the relation between his wage and estimated ability (past performance) is initially flat (zero). When the wage is positive, it equals one-half times the difference between the surplus that the board brings to the table and the surplus that the CEO brings. Whether the wage is increasing in estimated ability depends on whose surplus is increasing faster in $\hat{\alpha}$. Since the board's surplus is unbounded but the CEO's is bounded, it must be that the wage is eventually increasing in estimated ability. What we have not been able to establish is how the wage varies with estimated ability between $\tilde{A}$ and $\hat{A}$.[16] Overall, Proposition 7 predicts that the level at which the *non*contingent portion of a CEO's compensation is set should be insensitive to past performance for relatively low levels of past performance, but sensitive at relatively high levels of past performance.

It is worth noting that even if the CEO's wage is nonincreasing in his estimated ability, his overall well-being, $w + (p\Phi + 1 - p)b$, is increasing in his estimated ability.

III. Extensions

A. *The Board Has a Preference for the Incumbent CEO*

It is easy to imagine that the board has a preference for keeping the incumbent CEO. This could be a result of personal loyalty to the CEO—after all, many a directorship is the result of close ties between the CEO and the director (see, e.g., Mace, 1971). Alternatively, an incumbent CEO may take actions to entrench himself.

Let m be the additional value that an incumbent CEO yields the board. If $\hat{\alpha}$ is the board's estimate of his ability, then the board will treat him *as if* his estimated ability were $\hat{\alpha} + m \equiv \tilde{\alpha}$. It follows that the results from the previous section continue to hold, except with $\tilde{\alpha}$ replacing $\hat{\alpha}$. In particular, the next proposition is an immediate corollary of our earlier results.

PROPOSITION 8: *As the additional value that the incumbent CEO yields the board, m, rises the following occur:*

(*i*) *the intensity with which the current board monitors the CEO decreases;*
(*ii*) *the independence of the future board decreases; and*
(*iii*) *the minimum estimated ability for the incumbent CEO at which he will be retained prior to bargaining decreases.*

(The three results follow from Propositions 1, 6, and 2, respectively.)

In other words, Proposition 8 simply indicates that the more the board values the incumbent CEO independent of his ability, the less intensely he will be monitored by the board and the lower the standard to which he will held by the board. These results are consistent with the widely held belief that entrenched CEOs or CEOs who have cultivated personal loyalty are less scrutinized and face lower standards.

[16] We have been able to establish it, however, for specific examples. For instance, wage is strictly increasing in estimated ability (above $\tilde{A}$) if $d(p) = \frac{1}{2}p^2$, $\bar{k}_0 = 1$, $\tau = 1$, $\tau_0 = \frac{1}{2}$, $s = 1$, and $b = \frac{1}{4}$.

To the extent m is endogenous, Proposition 8 predicts that a CEO would undertake activities that raise m. An example of such an activity is given by Andrei Shleifer and Robert W. Vishny (1989). They argue that CEOs attempt to reduce the probability that they will be dismissed by making investments that are more profitable under them than any replacement CEO. Even if such investments decrease firm value, a CEO has an incentive to make them because they raise his value vis-à-vis a replacement.

Proposition 8 identifies another cost of entrenchment in addition to Shleifer and Vishny's investment-distortion cost: The more entrenched the CEO is, the less intensely he is monitored. Consequently, the board is less likely to identify problem CEOs who should be dismissed (even if the benefit m must be foregone), which further reduces expected firm value.

This analysis also shows that a CEO is better off with his friends on the board (i.e., people for whom m is positive). A CEO is, therefore, likely to use whatever influence he has to put directors who will be loyal to him on the board and to ensure the loyalty of those already on board.[17] Given this, it is not surprising that boards often become interlocked (see Hallock for evidence).

B. *Effects of Board Action on Share Price*

How is the value of the firm affected by the board's decision to fire or to retain the CEO? We consider the answer at two points: prior to when the board could obtain a signal and after it would have obtained a signal. Without loss of generality, we again assume the incumbent CEO offers no additional benefit to the board (i.e., $m = 0$).

Prior to monitoring, but after the first fire/retain decision, the value of the firm is[18]

$$P^*V + (1 - P^*)\hat{\alpha}. \tag{16}$$

Since P^* is decreasing in $\hat{\alpha}$ (recall Proposition 1), (16) need not be increasing in $\hat{\alpha}$. If, however, the disutility-of-effort function, $d(\cdot)$, is convex enough, then (16) will be increasing in $\hat{\alpha}$.

LEMMA 3: *If*

$$d(p) \le d(1) - \frac{d(1) - d(0)}{\log(2)} \log(2 - p) \quad \forall p \in [0, 1], \tag{17}$$

then (16) *is increasing in* $\hat{\alpha}$.

We will henceforth assume that (17) holds.[19]

Let FV equal (16) under an incumbent CEO and let FV_0 equal (16) under a replacement CEO. The probability of realizing a first-period profit such that the CEO is *dismissed* is

$$\Phi\left(\frac{\tau_0 + r}{r} A_c \sqrt{\frac{r\tau_0}{r + \tau_0}}\right),$$

where A_c is the cutoff ability level defined in Proposition 2. Consequently, at the beginning of the game the firm will be worth

$$\mathbb{E}_x\left\{\text{FV} \,\middle|\, x_1 \ge \frac{\tau_0 + r}{r} A_c\right\} \times \left(1 - \Phi\left(\frac{\tau_0 + r}{r} A_c \sqrt{\frac{r\tau_0}{r + \tau_0}}\right)\right) + \text{FV}_0 \Phi\left(\frac{\tau_0 + r}{r} A_c \sqrt{\frac{r\tau_0}{r + \tau_0}}\right).$$

[17] For an extreme example see Bryan Burrough and John Helyar's (1990) discussion of the board of RJR-Nabisco.

[18] Our analysis ignores the present discounted value of the firm *beyond* the period considered by our model. This is slightly problematic because, as we argued in Propositions 4 and 6, we should expect hysteresis across CEO regimes. Given, however, the relatively long tenure of CEOs (ten years on average—see, e.g., Hermalin and Weisbach, 1988), this future omitted part of firm value will generally represent a very small portion of the firm's value.

[19] If $d(\cdot)$ is not convex enough, then (16) may be decreasing in $\hat{\alpha}$ over some range of $\hat{\alpha}$. To see this, suppose $d(\cdot)$ were affine. Then there would exist an $\hat{\alpha}^*$ such that $P^* = 1$, for $\hat{\alpha} < \hat{\alpha}^*$ and $P^* = 0$, for $\hat{\alpha} > \hat{\alpha}^*$. Since $V > \hat{\alpha}$, this would imply that (16) must decrease as $\hat{\alpha}$ crosses $\hat{\alpha}^*$.

After the first realization of profits, x_1, the firm's value is

$$x_1 + \chi \text{FV} + (1 - \chi)\text{FV}_0, \tag{18}$$

where

$$\chi = \begin{cases} 1 & \text{if } P^*V + (1 - P^*)\hat{\alpha} - \bar{k}_0 d(P^*) \\ & > U_0 \\ 0 & \text{otherwise} \end{cases}$$

indicates whether the incumbent CEO is retained or fired.

From Section II, subsection D,

$$\text{FV}_{\hat{\alpha}=A_c} - \bar{k}_0 d(P^*_{\hat{\alpha}=A_c})$$

$$= \text{FV}_0 - \bar{k}_0 d(P_0) \equiv U_0.$$

Moreover, from Proposition 1, $P^*_{\hat{\alpha}=A_c} < P_0$;[20] hence, $\text{FV}_{\hat{\alpha}=A_c} < \text{FV}_0$. There thus exists an interval of $\hat{\alpha}$'s starting at A_c such that investors would prefer that the CEO be fired, but such that the board would prefer to retain the CEO. Consequently, there is a discontinuous drop in the value of the firm at $\hat{\alpha} = A_c$. Since $\hat{\alpha}$ and x_1 are monotonically related, we can conclude that firm value, expression (18), is increasing for all values of x_1 *except* at

$$x_1 = \frac{\tau_0 + r}{r} A_c,$$

where there is a discontinuous drop in firm value.

We summarize the analysis so far.

PROPOSITION 9: *Higher first-stage profit is positively related to whether the CEO keeps his job. But higher firm value is not monotonically related to whether the CEO keeps his job. Moreover, there is a range of first-stage profits such that investors would prefer that the CEO be fired, while the board prefers to retain him.*

Remark 3: Proposition 9 suggests that earnings (i.e., x_1) should be a better predictor of CEO turnover than share price, which is consistent with the empirical literature (see, e.g., Weisbach or Murphy and Zimmerman).

Proposition 9 and the discussion preceding it indicates that a tension can exist between investors and directors over whether the CEO should be fired, with the investors preferring to fire and the directors preferring to retain. This provides an explanation for the common phenomenon of investors seeming more eager than the board to dismiss management. It can also explain why takeovers and other costly control contests can be worth mounting.

Now we turn to the stock reaction when the board bases its fire/retain decision on its private signal. If the board fires the CEO, then the expected value of *future* cash flows is zero. Prior to evaluation, the expected value of future cash flows is positive. It follows, therefore, that:

PROPOSITION 10: *The firm's stock price falls if the CEO is fired on the basis of the board's private information.*

Finally, suppose $\bar{k}$ is *not* known to investors. Ignorance of $\bar{k}$ does not change Proposition 10, so we will focus on what happens when the board fires the CEO based on public information. For any value of $\bar{k}$ if the board wants to fire the CEO, investors would also want the CEO fired. If we imagine a distribution over $\bar{k}$ such that it is uncertain whether the board will fire the CEO for a given level of first-stage profits, then firing the CEO will be considered good news by investors and will cause the stock price to rise. The stock price will also rise because firing the CEO signals that the board is relatively more independent than was anticipated and will, thus, monitor more intensely. This yields:

PROPOSITION 11: *Suppose that the board's independence is unknown to investors. Then the firm's stock price rises if the CEO is fired on the basis of public information.*

[20] Recall P_0 is the optimal intensity of monitoring for a replacement CEO.

Remark 4: Our result that the stock price reaction to a CEO dismissal differs depending on whether the board used public or private information is consistent with, and could even explain, the ambiguous relation between CEO dismissal and stock price reaction found in event studies of CEO turnover. See Warner et al. for a survey of these studies and a similar explanation for their inconclusive findings.

IV. A Nonbargaining Interpretation

A potential concern is the realism of the bargaining game and the extent to which our results depend on it. Hence, it is worth reinterpreting the model in a way that does not depend on bargaining.

In this interpretation, the timing is the same, except that the bargaining stage, stage 4, no longer exists. We also want to reinterpret stage 5: Let

$$p(V - \hat{\alpha}) + \hat{\alpha} - \bar{k}d(p)$$

be the firm's expected profit; where, now, $\bar{k}d(p)$ is the cost of monitoring. We now interpret $\bar{k}$ to be a cost parameter known to the firm's decision makers, but possibly unknown to investors.

We assume, now, that a board's level of monitoring, p, is an intrinsic attribute of the board. In particular, it is invariant with respect to $\hat{\alpha}$ or other parameters. Boards that monitor more—have a higher p—are more costly for the firm than boards that monitor less—have a lower p. That is, $d'(p) > 0$. We assume, additionally, that this *marginal* cost is also increasing in p (i.e., $d''(p) > 0$).[21]

We can now reinterpret Proposition 1 as a statement about board composition and their underlying parameters.

COROLLARY 1: *Under the alternative interpretation of this section, the level of board independence, p, is*

(i) *decreasing with the prior estimate of the CEO's ability, $\hat{\alpha}$;*
(ii) *decreasing with the precision of the prior estimate, τ; but*
(iii) *increasing with the precision of the signal, s.*

Observe that Corollary 1(i) and (ii) are substitutes for Proposition 4 (surviving CEOs have easier boards), while Corollary 1(i) is a substitute for Proposition 6 (board independence is decreasing with estimated CEO ability).

Propositions 2 and 3 continue to hold under this alternative interpretation (although Proposition 3 has little economic meaning). The loyalty-entrenchment result, Proposition 8, continues to hold under this alternative interpretation. Likewise, provided $\bar{k}$ is unknown to investors, the share-price results, Propositions 9–11, also continue to hold under this alternative interpretation.

Most of our results are, therefore, not dependent on the existence of a bargaining stage. Rather they are driven by combining a matching model, similar to Jovanovic's, with endogenous monitoring. Bargaining enables us to address the central enigma, set forth by Berle and Means and others, of how a seemingly inefficient institution has survived. In particular, it serves to explain how, why, and when CEOs have a say over who serves on the board. It also serves to explain how, despite this say, the board can still provide a valuable monitoring role.

V. Policy Implications of the Model

As corporate governance has remained essentially the same since the days of Berle and Means, so too have the criticisms and proposed reforms of it. For example, Lipton and Lorsch call for a number of changes, including a smaller board (to reduce free-riding),

[21] For example, there could be benefits to having inside directors on the board (perhaps to groom them as potential successors to the CEO—see Richard F. Vancil [1987] on this point) or less-intensely-monitoring outside directors (perhaps because they bring needed expertise to the firm). These benefits are increasing in the number of such directors, but with diminishing marginal returns. When these directors are seen as the opportunity cost of more-intensely-monitoring outside directors, we have $d'(p) > 0$ and $d''(p) > 0$.

more outsiders, more meetings, director pay linked to stock performance, and appointment of a "lead" director (if not the chairman) who is separate from the CEO.[22] These policies, in Lipton and Lorsch's view, would lead to better monitoring of the CEO.

In the context of our model, each element of the Lipton and Lorsch proposal serves to reduce the $\bar{k}$ of the board. For example, stock-based incentives lower the k_i's of all directors, while replacing high-k insiders with lower-k outsiders lowers average k. At first glance, one might think that efforts to lower the board's $\bar{k}$ through regulation or political pressure would lead to more effective monitoring.

This argument, however, ignores the nature of the equilibrium in the model. The CEO and board bargain over the *effective* $\bar{k}$, which takes account of all incentives that potential directors will have while they are on the board. As long as the bargaining process is itself unaffected by reforms, the equilibrium $\bar{k}$ will be little affected.

It follows that we need to distinguish between those policies that will affect the bargaining process and those that will not. For instance, requiring a specified fraction of the board to be outsiders would result in an outsider-dominated board, but not necessarily one that is more independent than the insider-dominated board that would otherwise prevail—the CEO and board members will have latitude in the selection process to offset whatever benefits are created by exogenously imposed "independence."[23] On the other hand, the model suggests that requiring incentive pay for directors could have an effect: By lowering $\bar{k}_0$, this requirement would affect the bargaining, leading to more independent boards and greater monitoring (see Proposition 6). Moreover, because of hysteresis, these benefits can persist—although entropy could lead them to diminish over time.

Of course this analysis begs the question of why corporations do not voluntarily adopt effective reforms such as this. One answer is that just as the board and the CEO negotiate over board composition, they would also negotiate over the implementation of reforms. Provided his past success gave him sufficient bargaining power, the CEO would be able to block or blunt such reforms.[24]

VI. Concluding Remarks

A recent *Harvard Business Review* "Perspectives Section" provides some insight into the realism of our model (Smale et al.). John Smale, who became the nonexecutive chairman of General Motors following Robert Stempel's forced resignation, describes policies adopted by the GM board that have dramatically improved its effectiveness. In contrast, Alan Patricof, a leading venture capitalist, argues: "Deep down [CEOs] really wish they didn't have boards. That's why, at the end of the day, most independent directors get neutralized in one fashion or another (Smale et al., p. 8)." A model of corporate governance should be consistent with both perspectives; it should explain both how some boards are active monitors of management, yet how some CEOs are able to avoid scrutiny.

By studying the determinants of board composition as a bargaining process, our model is consistent with both active monitoring in some firms and CEO dominance in others. The process by which GM acquired a strong board is illustrative of the model's logic: The company had a crisis induced by poor profits and the

[22] This proposal is fairly representative of the many governance reforms that have been proposed by the business press, academics, and business people themselves (see American Law Institute, 1982; Jensen, 1993; John G. Smale et al., 1995). See, too, Brickley et al. (1997) for a discussion and evidence on one of these potential reforms, the separation of the CEO and chairman's positions.

[23] Since an outsider is simply someone with no other ties to the corporation, it is hard to imagine that high-k outsiders could not be found (e.g., through interlocking boards).

[24] An additional prediction of our analysis is that new firms or those with weak CEOs (i.e., firms less subject to the agency problems considered here) will be the first to adopt reforms.

board was forced to act. The new CEO had no bargaining power and, thus, had to contend with an active board. None of this would have happened had the previous managers performed better; they would have maintained their jobs and their control over the board. Subsequently, after a period of good performance, GM went back to a more traditional arrangement of have the CEO also serve as chairman (*Wall Street Journal,* December 5, 1995 p. B1).

The model is consistent with a number of empirical regularities: CEO turnover is negatively related to performance and this relation is stronger when the board is more independent. The probability that independent directors are added to the board increases following poor corporate performance. And boards tend to become less independent over the course of the CEO's career. The model also explains why management turnover is more related to earnings than to stock returns. Finally, the model provides insight into the effectiveness of various policies designed to enhance the board's monitoring.

Despite the model's consistency with existing empirical evidence, a number of directions for future research remain. One is to model the board's operation in greater detail. For instance, we have assumed that the board chooses a common intensity of monitoring, p. What we have not considered is how the board implements this choice. For instance, does p represent the collective output of the board (e.g., what it does at board meetings) or is it an aggregate of individual directors' efforts (e.g., carefully reading reports prior to board meetings)? If it is the second, to what extent is the board able to overcome the problem of free-riding endemic to team production (see, e.g., Holmstrom, 1982)? Once free-riding among directors is an issue, the dynamics of board composition become more complicated. For example, the continuing directors can reduce their own workloads by adding very independent directors (i.e., low-k directors) to the board. This, in turn, could offset the entropy prediction of Proposition 4.

For example, unlike most American companies, a German or Japanese company typically has strong ties to one particular bank and representatives of this bank usually serve on the company's board.[25] These representatives presumably have a strong interest in the company's well-being.[26] The diligence of the rest of the board is unclear; free-riding considerations would tend to reduce their effectiveness, while the bank representatives have incentives to ensure that directors be selected who will be less likely to free ride. Similar issues could be expected to arise in family-owned firms in the United States.

One limitation of our model is that it focuses solely on the monitoring role of boards. The institutional literature (see, e.g., Mace or Vancil) emphasizes that boards also play important roles providing information and advice to management, and serving as a training ground for future CEOs. A richer model of boards should take into account these roles as well. To the extent they represent opportunity costs of monitoring [make $d'(p) > 0$], these other roles complement our analysis.

Our model could also be extended to investigate the transition from an entrepreneurial firm to a managerial firm. In this transition, an entrepreneur (or his or her venture capital firm) has an incentive to maximize the value of the firm by minimizing the impact of the entropy problem.

A last topic for future research would be to consider noncorporate situations where boards play a monitoring role. For example, universities, trusts, and other nonprofit institutions all have bodies that function much like corporate boards of directors. Much of the analysis presented above would seem equally applicable to these boards, but as with international comparisons further work is worth pursuing.[27]

[25] See Kaplan (1994a, b) for recent evidence on the effects of these banking relationships on corporate governance in Germany and Japan.

[26] Although it should be remembered that such directors are themselves agents (of the bank), which could create a second set of agency problems. Despite this, it is still reasonable to expect these directors to be more concerned about the firm's profits than other directors.

[27] See William G. Bowen (1994) for a discussion of the differences between profit and nonprofit boards.

APPENDIX: PROOFS

PROOF OF LEMMA 1:

$$\frac{\partial V}{\partial \hat{\alpha}} = \Phi + \hat{\alpha}\phi\sqrt{H}$$

$$+ \frac{H}{\tau}(Y_c - \hat{\alpha})\phi\sqrt{H}\left(\text{note } \frac{\partial V}{\partial Y_c} = 0\right)$$

$$= \Phi > 0 \left(\text{recall } H = \frac{s\tau}{\tau + s} \quad \text{and}\right.$$

$$\left. Y_c - \hat{\alpha} = -\frac{\tau + s}{s}\hat{\alpha}\right).$$

Consequently,

$$\frac{\partial(V - \hat{\alpha})}{\partial \hat{\alpha}} = \Phi - 1 < 0.$$

PROOF OF PROPOSITION 1:

Let Ω be the expression to be maximized in (7). Consider (i), if $\hat{\alpha} \geq 0$, then:

$$\frac{\partial^2 \Omega}{\partial \hat{\alpha} \partial p} = \frac{\partial[V - \hat{\alpha}]}{\partial \hat{\alpha}} < 0 \text{ by Lemma 1},$$

so, by the usual comparative statics, $\partial P^*/\partial \hat{\alpha} < 0$. Similarly,

$$\frac{\partial^2 \Omega}{\partial \bar{k} \partial p} = -d'(p) < 0; \text{ and } \frac{\partial^2 \Omega}{\partial \tau \partial p} = \frac{\partial V}{\partial \tau}$$

$$= \left(-1 + \frac{1}{2}\frac{s}{s + \tau}\right)\frac{\sqrt{H}}{\tau^2}\phi < 0$$

(where the second result uses the fact that $\partial V/\partial Y_c = 0$). Hence, $\partial P^*/\partial \bar{k} < 0$ and $\partial P^*/\partial \tau < 0$. Finally,

$$\frac{\partial^2 \Omega}{\partial s \partial p} = \frac{\partial V}{\partial s} = \frac{\tau}{2(\tau + s)^2\sqrt{H}}\phi > 0,$$

so, by the usual comparative statics, $\partial P/\partial s > 0$.

PROOF OF LEMMA 2:

Consider bargaining with a new CEO. If this bargaining is unsuccessful, the board can hire yet another CEO. Hence, from (12), bargaining entails maximizing

$$(P^*V_0 - \bar{k}_0 d(P^*) - w - U_0)$$
$$\times ((P^*\Phi + 1 - P^*)b + w)$$

with respect to $\bar{k}$ and w. Given the monotonic relationship between P^* and $\bar{k}$ (Proposition 1) we can equivalently maximize this product in P^* and w. The first-order conditions are

$$\text{(A1)} \quad (V_0 - \bar{k}_0 d'(P^*)) \times ((P^*\Phi + 1 - P^*)b + w) + (P^*V_0 - \bar{k}_0 d(P^*) - w - U_0) \times (\Phi - 1)b = 0 \quad \text{and}$$

$$\text{(A2)} \quad P^*V_0 - \bar{k}_0 d(P^*) - w - U_0 - ((P^*\Phi + 1 - P^*)b + w) \leq 0.$$

In equilibrium, $P^*V_0 - \bar{k}_0 d(P^*) - w = U_0$—one new CEO yields the board the same utility as another new CEO. It follows, then, from (A2) that $w = 0$. Plugging that back into (A1) yields

$$V_0 - \bar{k}_0 d'(P^*) = 0.$$

This is the first-order condition (8). Since P^* is monotonic in $\bar{k}$, the solution to the Nash bargaining game is therefore $\bar{k}_1 = \bar{k}_0$.

PROOF OF PROPOSITION 2:

Let U^E equal the equilibrium expected utility of the board if it bargains with the incumbent CEO. Since $P^*_{\bar{k}=\bar{k}_0}$ is the board's most preferred level of monitoring, we know

$$\text{(A3)} \quad U^E \leq P^*_{\bar{k}=\bar{k}_0}V + (1 - P^*_{\bar{k}=\bar{k}_0}) \times \max\{0, \hat{\alpha}\} - \bar{k}_0 d(P^*_{\bar{k}=\bar{k}_0}).$$

Using the envelope theorem, it is readily shown that the right-hand side of (A3) is increasing in $\hat{\alpha}$. Moreover, as $\hat{\alpha}$ goes to infinity, the right-hand side of (A3) also goes to infinity. Differentiating the right-hand side of (A3) with respect to τ using the envelope theorem yields

$$\text{(A4)}\qquad P^*_{\bar{k}=\bar{k}_0}\left[-1+\frac{1}{2}\frac{s}{\tau+s}\right]\frac{\sqrt{H}}{\tau^2}\phi<0$$

(the option value is decreasing with the precision with which the CEO's ability is estimated). Suppose, now, that $\hat{\alpha}=0$. It follows from (4) and (A4) that

$$\text{(A5)}\qquad P^*_{\bar{k}=\bar{k}_0}V-\bar{k}_0 d(P^*_{\bar{k}=\bar{k}_0}) < P_0V_0-\bar{k}_0 d(P_0)$$

$$\text{(A6)}\qquad = U_0.$$

Combining (A3) and (A6) establishes that $A_c>0$. Since the right-hand side of (A3) is continuous and increasing without bound but is less than U_0 for an estimated ability of 0, it follows that A_c exists and is unique.

PROOF OF PROPOSITION 3:

In the proof of Proposition 2, we established that

$$\text{(A7)}\qquad P^*_{\bar{k}=\bar{k}_0}(V_{\hat{\alpha}=A_c}-A_c)+A_c - \bar{k}_0 d(P^*_{\bar{k}=\bar{k}_0})-U_0=0$$

for all $\bar{k}_0$. Since (A7) holds for all $\bar{k}_0$, it is an identity. Differentiating (A7) with respect to $\bar{k}_0$ using the envelope theorem yields

$$\text{(A8)}\qquad [P^*\Phi+1-P^*]\frac{\partial A_c}{\partial \bar{k}_0} - d(P^*_{\bar{k}=\bar{k}_0})+d(P_0)=0.$$

From Proposition 1, $P^*_{\bar{k}=\bar{k}_0}<P_0$. Hence, since $d(\cdot)$ is an increasing function, it follows from (A8) that $\partial A_c/\partial \bar{k}_0<0$.

PROOF OF PROPOSITION 6:

There are two cases to consider: (i) the limited-liability constraint is binding ($w=0$) and (ii) it is not binding. Begin with case (i). From Proposition 1, the probability that the CEO is monitored is monotonic in $\bar{k}$, so maximizing (12) with respect $\bar{k}$ is equivalent to maximizing

$$\text{(A9)}\qquad (pV(\hat{\alpha},\tau)+(1-p)\hat{\alpha}-\bar{k}_0 d(p)-U_0) \times(p\Phi+1-p)$$

with respect to p (since the CEO will be retained, we know $\hat{\alpha}>0$). Define Ψ to equal (A9). By well-known comparative statics results, it is sufficient to show $\partial^2\Psi/\partial\hat{\alpha}\partial p$ is negative:

$$\text{(A10)}\qquad \frac{\partial^2\Psi}{\partial\hat{\alpha}\partial p}=2p(\Phi-1)^2+2(\Phi-1) +\frac{\partial\Phi}{\partial\hat{\alpha}}p[V-\hat{\alpha}-\bar{k}_0 d'(p)] +\frac{\partial\Phi}{\partial\hat{\alpha}}[p(V-\hat{\alpha})+\hat{\alpha} -\bar{k}_0 d(p)-U_0].$$

Using the first-order condition for (A9), (A10) can be rewritten as

$$\frac{\partial^2\Psi}{\partial\hat{\alpha}\partial p}=2p(\Phi-1)^2+2(\Phi-1) +\frac{\partial\Phi}{\partial\hat{\alpha}}\frac{V-\hat{\alpha}-\bar{k}_0 d'(p)}{1-\Phi}.$$

Hence, we have

$$\text{(A11)}\qquad \frac{\partial^2\Psi}{\partial\hat{\alpha}\partial p}<2(\Phi-1)^2+2(\Phi-1) +\frac{\partial\Phi}{\partial\hat{\alpha}}\frac{V-\hat{\alpha}}{1-\Phi}.$$

We will now show that the right-hand side of (A11) is negative. Note, first, that

$$\frac{\partial \Phi}{\partial \hat{\alpha}} = \sqrt{H}\left(\frac{\tau + s}{s}\right)\phi\left(\sqrt{H}\left(\frac{\tau + s}{s}\right)\hat{\alpha}\right).$$

Next, make the substitution

$$z = \sqrt{H}\left(\frac{\tau + s}{s}\right)\hat{\alpha} \tag{A12}$$

and use (6) to rewrite the right-hand side of (A11) as

$$2(\Phi(z) - 1)^2 + 2(\Phi(z) - 1) + \sqrt{H}\left(\frac{\tau + s}{s}\right)\phi(z)\left(-\frac{zs}{(\tau + s)\sqrt{H}} + \frac{\phi(z)\sqrt{H}}{(1 - \Phi(z))\tau}\right).$$

Simplifying, this reduces to

$$2(\Phi(z) - 1)^2 + 2(\Phi(z) - 1) - z\phi(z) + \frac{\phi(z)^2}{1 - \Phi(z)}. \tag{A13}$$

Straightforward calculations reveal that this last expression is negative for all $z \geq 0$ (and, hence, for all $\hat{\alpha} \geq 0$). So the right-hand side of (A11) is negative, which, from (A11), entails

$$\frac{\partial^2 \Psi}{\partial \hat{\alpha} \partial p} < 0.$$

Now consider case (ii). From Proposition 5, the level of monitoring satisfies (10):

$$V - \hat{\alpha} - \bar{k}_0 d'(p) - (1 - \Phi)b = 0. \tag{A14}$$

Let Ω be the left-hand side of (A14). The result follows if $\partial\Omega/\partial\hat{\alpha} < 0$.

$$\begin{aligned} \frac{\partial \Omega}{\partial \hat{\alpha}} &= \Phi - 1 + b\frac{\partial \Phi}{\partial \hat{\alpha}} \\ &= \Phi - 1 + \frac{\partial \Phi}{\partial \hat{\alpha}}\frac{V - \hat{\alpha} - \bar{k}_0 d'(p)}{1 - \Phi} \qquad \text{(A15)} \\ &< \Phi - 1 + \frac{\partial \Phi}{\partial \hat{\alpha}}\frac{V - \hat{\alpha}}{1 - \Phi} \end{aligned}$$

where (A15) follows from (A14). Using the z-transformation, we have

$$\frac{\partial \Omega}{\partial \hat{\alpha}} < \Phi(z) - 1 - z\phi(z) + \frac{\phi(z)^2}{1 - \Phi(z)}.$$

Straightforward calculations reveal that this last expression is negative for all $z \geq 0$ (and, hence, for all $\hat{\alpha} \geq 0$).

PROOF OF PROPOSITION 7:

Consider $\tilde{A}$ first. From (14), $w = 0$ if

$$[p(V - \hat{\alpha}) + \hat{\alpha} - \bar{k}_0 d(p) - U_0] - [p\Phi b + (1 - p)b] < 0.$$

At $\hat{\alpha} = A_c$, the first bracketed term is zero, while the second is strictly positive. The existence of $\tilde{A}$ then follows from continuity.

Turn to $\hat{A}$. If $w > 0$, then

$$2w = [p(V - \hat{\alpha}) + \hat{\alpha} - \bar{k}_0 d(p) - U_0] - [p\Phi b + (1 - p)b].$$

The first bracketed term increases without limit in $\hat{\alpha}$, while the second has an upper bound of b. It follows then that beyond a certain level, $\hat{A}$, that w must be increasing in $\hat{\alpha}$.

PROOF OF LEMMA 3:

From Proposition 2, we know $\hat{\alpha} \geq A_c > 0$ or $\hat{\alpha} = 0$. Differentiate (16) with respect to $\hat{\alpha}$:

$$1 + P^*(\Phi - 1) + (V - \hat{\alpha})\frac{dP^*}{d\hat{\alpha}}.$$

From (8), this can be rewritten as

$$1 + P^*(\Phi - 1) + \bar{k}d'(P^*)\frac{dP^*}{d\hat{\alpha}}.$$

Also from (8), it is readily shown that

$$\frac{dP^*}{d\hat{\alpha}} = \frac{\Phi - 1}{\bar{k}d''(P^*)}.$$

So (16) is increasing in $\hat{\alpha}$ if

$$\text{(A16)} \quad 1 + (\Phi - 1)\left(P^* + \frac{d'(P^*)}{d''(P^*)}\right) \geq 0.$$

Since $\hat{\alpha} \geq 0$, $\Phi \geq 1/2$, so (A16) holds if

$$\text{(A17)} \qquad 2 \geq P^* + \frac{d'(P^*)}{d''(P^*)}.$$

Define $\bar{d}(p)$ to equal the right-hand side of (17). It is readily shown that

$$p + \frac{\bar{d}'(p)}{\bar{d}''(p)} = 2,$$

so (A17) follows if $d'(p)/d''(p) \leq \bar{d}'(p)/\bar{d}''(p)$. By (17), $d(\cdot)$ is more convex than $\bar{d}(\cdot)$. By adapting a well-known result on when one individual will have a greater Arrow-Pratt measure of absolute risk aversion than another individual (see, e.g., Chi-fu Huang and Robert H. Litzenberger, 1988 p. 29), it readily follows that $d'(p)/d''(p) \leq \bar{d}'(p)/\bar{d}''(p)$ for all $p \in [0, 1]$.

REFERENCES

Aghion, Philippe and Bolton, Patrick. "Contracts as a Barrier to Entry." *American Economic Review*, June 1987, *77*(3), pp. 388–401.

Agrawal, Anup and Knoeber, Charles R. "Firm Performance and Mechanisms to Control Agency Problems between Managers and Shareholders." *Journal of Financial and Quantitative Analysis*, September 1996, *31*(3), pp. 377–97.

American Law Institute. *Principles of corporate governance and structure: Restatement and recommendations*. American Law Institute, Tenetative Draft No. 1, 1982.

Baysinger, Barry D. and Butler, Henry N. "Corporate Governance and the Board of Directors: Performance Effects of Changes in Board Composition." *Journal of Law, Economics and Organization*, Spring 1985, *1*(1), pp. 101–24.

Berle, Adolph A. and Means, Gardiner C. *The modern corporation and private property*. New York: MacMillan, 1932.

Bhagat, Sanjai and Black, Bernard. "Do Independent Directors Matter?" Working paper, Columbia University, 1996.

Booth, James R. and Deli, Daniel N. "Factors Affecting the Number of Outside Directorships Held by CEOs." *Journal of Financial Economics*, January 1996, *40*(1), pp. 81–104.

Borokhovich, Kenneth A.; Parrino, Robert P. and Trapani, Teresa. "Outside Directors and CEO Selection." *Journal of Financial and Quantitative Analysis*, September 1996, *31*(3), pp. 337–55.

Bowen, William G. *Inside the boardroom: Governance by directors and trustees*. New York: Wiley, 1994.

Brickley, James A.; Coles, Jeffrey L. and Jarrell, Gregg A. "The Leadership Structure of the Corporation: Separating the Positions of CEO and Chairman of the Board." *Journal of Corporate Finance*, June 1977, *3*(3), pp. 189–220.

Brickley, James A.; Coles, Jeffrey L. and Terry, Rory L. "Outside Directors and the Adoption of Poison Pills." *Journal of Financial Economics*, June 1994a, *35*(3), pp. 371–90.

______. "Appendix to Accompany 'Outside Directors and the Adoption of Poison Pills'." Working paper, Arizona State University, 1994b.

Brickley, James A. and James, Christopher M. "The Takeover Market, Corporate Board Composition, and Ownership Structure: The Case of Banking." *Journal of Law and Economics*, April 1987, *30*(1), pp. 161–80.

Brown, William and Maloney, Michael T. "Acquisition Performance and Corporate Board Composition." Working paper, Clemson University, 1992.

Burrough, Bryan and Helyar, John. *Barbarians at the gate*. New York: Harper and Row, 1990.

Byrd, John W. and Hickman, Kent A. "Do Outside Directors Monitor Managers? Evidence from Tender Offer Bids." *Journal of Fi*

Financial Economics, October 1992, *32*(2), pp. 195–221.

Cotter, James F.; Shivdasani, Anil and Zenner, Marc. "Do Independent Directors Enhance Target Shareholder Wealth during Tender Offers?" *Journal of Financial Economics*, February 1997, *43*(2), pp. 195–218.

Coughlan, Anne T. and Schmidt, Ronald M. "Executive Compensation, Managerial Turnover, and Firm Performance: An Empirical Investigation." *Journal of Accounting and Economics*, April 1985, *7*(1/2/3), pp. 43–66.

DeAngelo, Harry and DeAngelo, Linda. "Proxy Contests and the Governance of Publicly Held Corporations." *Journal of Financial Economics*, June 1989, *23*(1), pp. 29–60.

DeGroot, Morris H. *Optimal statistical decisions*. New York: McGraw-Hill, 1970.

Demb, Ada and Neubauer, F.-Friedrich. *The corporate board*. Oxford: Oxford University Press, 1992.

Fama, Eugene F. "Agency Problems and the Theory of the Firm." *Journal of Political Economy*, April 1980, *88*(2), pp. 288–307.

Fama, Eugene F. and Jensen, Michael C. "Separation of Ownership and Control." *Journal of Law and Economics*, June 1983, *26*(2), pp. 301–26.

Gertner, Robert and Kaplan, Steven N. "The Value-Maximizing Board." Working paper, University of Chicago, 1997.

Hallock, Kevin. "Reciprocally Interlocking Boards of Directors and Executive Compensation." *Journal of Financial and Quantitative Analysis*, September 1997, *32*(3), pp. 331–44.

Hermalin, Benjamin E. and Weisbach, Michael S. "The Determinants of Board Composition." *Rand Journal of Economics*, Winter 1988, *19*(4), pp. 589–606.

———. "The Effects of Board Composition and Direct Incentives on Firm Performance." *Financial Management*, Winter 1991, *20*(4), pp. 101–12.

Hirshleifer, David and Thakor, Anjan V. "Managerial Performance, Boards of Directors, and Takeover Bidding." *Journal of Corporate Finance*, March 1994, *1*(1), pp. 63–90.

Holmstrom, Bengt. "Moral Hazard in Teams." *Bell Journal of Economics*, Autumn 1982, *13*(2), pp. 324–40.

———. "Managerial Incentive Problems—A Dynamic Perspective." Working paper, Northwestern University, 1993.

Huang, Chi-fu and Litzenberger, Robert H. *Foundations for financial economics*. Amsterdam: North-Holland, 1988.

Jensen, Michael C. "The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems." *Journal of Finance*, July 1993, *48*(3), pp. 831–80.

Jensen, Michael C. and Murphy, Kevin J. "Performance Pay and Top-Management Incentives." *Journal of Political Economy*, April 1990, *98*(2), pp. 225–64.

Jovanovic, Boyan. "Job Matching and the Theory of Turnover." *Journal of Political Economy*, October 1979, *87*(5), pp. 972–90.

Kaplan, Steven N. "Top Executives, Turnover, and Firm Performance in Germany." *Journal of Law, Economics and Organization*, April 1994a, *10*(1), pp. 142–59.

———. "Top Executive Rewards and Firm Performance: A Comparison of Japan and the United States." *Journal of Political Economy*, June 1994b, *102*(3), pp. 510–46.

Kaplan, Steven N. and Reishus, David. "Outside Directorships and Corporate Performance." *Journal of Financial Economics*, October 1990, *27*(2), pp. 389–410.

Kini, Omesh; Krakaw, William and Mian, Shehzad. "Corporate Takeovers, Firm Performance, and Board Composition." *Journal of Corporate Finance*, April 1994, *1*(3/4), pp. 383–412.

Klein, April. "Firm Productivity and Board Committee Structure." *Journal of Law and Economics*, 1998 (forthcoming).

Lipton, Martin and Lorsch, Jay W. "A Modest Proposal for Improved Corporate Governance." *Business Lawyer*, November 1992, *48*(1), pp. 59–77.

Lorsch, Jay W. and MacIver, Elizabeth. *Pawns or potentates: The reality of America's corporate boards*. Boston: Harvard Business School Press, 1989.

MacAvoy, Paul; Cantor, Scott, Jr.; Dana, James D. and Peck, Sarah. "ALI Proposals for Increased Control of the Corporation by the Board of

Directors: An Economic Analysis.'' *Statement of the business roundtable on the American Law Institute's proposed principles of corporate governance and structure: Restatement and recommendations*. New York: Business Roundtable, 1983.

Mace, Myles L. *Directors: Myth and reality*. Boston: Harvard Business School Press, 1971.

Mayers, David; Shivdasani, Anil and Smith, Clifford. ''Board Composition in the Life Insurance Industry.'' *Journal of Business*, January 1997, *70*(1), pp. 33–62.

Murphy, Kevin J. and Zimmerman, Jerold L. ''Financial Performance Surrounding CEO Turnover.'' *Journal of Accounting and Economics*, January/April/July 1993, *16*(1/2/3), pp. 273–315.

Noe, Thomas H. and Rebello, Michael. ''The Design of Corporate Boards: Composition, Factions, and Turnover.'' Working paper, Georgia State University, 1996.

Roe, Mark. *Strong managers, weak owners: The political roots of American corporate finance*. Princeton, NJ: Princeton University Press, 1994.

Rosenstein, Stuart and Wyatt, Jeffrey G. ''Outside Directors, Shareholder Independence, and Shareholder Wealth.'' *Journal of Financial Economics*, August 1990, *26*(2), pp. 175–92.

Shivdasani, Anil. ''Board Composition, Ownership Structure, and Hostile Takeovers.'' *Journal of Accounting and Economics*, January/April/July 1993, *16*(1/2/3), pp. 167–98.

Shleifer, Andrei and Vishny, Robert W. ''Management Entrenchment: The Case of Manager-Specific Investments.'' *Journal of Financial Economics*, November 1989, *25*(1), pp. 123–39.

Smale, John G.; Patricof, Alan J.; Henderson, Denys; Marcus, Bernard and Johnson, David W. ''Redraw the Line between the Board and the CEO.'' *Harvard Business Review*, March/April 1995, *73*(2), pp. 5–12.

Stigler, George J. and Friedland, Claire. ''The Literature of Economics: The Case of Berle and Means.'' *Journal of Law and Economics*, June 1983, *26*(2), pp. 237–68.

Vancil, Richard F. *Passing the baton: Managing the process of CEO succession*. Boston: Harvard Business School Press, 1987.

Wall St. Journal**.** ''GM Decides One Head is Better Than Two.'' December 5, 1995, p. B1.

Warner, Jerold B.; Watts, Ross L. and Wruck, Karen H. ''Stock Prices and Top Management Changes.'' *Journal of Financial Economics*, January/March 1988, *20*(1/2), pp. 461–92.

Warther, Vincent A. ''Board Effectiveness and Board Dissent: A Model of the Board's Relationship to Management and Shareholders.'' Working paper, University of Michigan, 1994.

Weisbach, Michael S. ''Outside Directors and CEO Turnover.'' *Journal of Financial Economics*, January/March 1988, *20*(1/2), pp. 431–60.

Yermack, David. ''Higher Valuation of Companies with a Small Board of Directors.'' *Journal of Financial Economics*, February 1996, *40*(2), pp. 185–212.

The Simple Economics of Easter Island: A Ricardo-Malthus Model of Renewable Resource Use

By JAMES A. BRANDER AND M. SCOTT TAYLOR*

This paper presents a general equilibrium model of renewable resource and population dynamics related to the Lotka-Volterra predator-prey model, with man as the predator and the resource base as the prey. We apply the model to the rise and fall of Easter Island, showing that plausible parameter values generate a "feast and famine" pattern of cyclical adjustment in population and resource stocks. Near-monotonic adjustment arises for higher values of a resource regeneration parameter, as might apply elsewhere in Polynesia. We also describe other civilizations that might have declined because of population overshooting and endogenous resource degradation. (*JEL* Q20, N57, J10)

The world of the late twentieth century is much more heavily populated and has much higher average living standards than any previous period in human history. However, widely publicized concerns have been expressed over whether per capita real income can continue to increase or even be maintained at current levels in the face of rapid population growth and environmental degradation. Economists normally tend to be skeptical about such claims, largely on the basis of the historical record. At various times in the past, people have worried about overpopulation and environmental degradation, yet the past, at least as we imagine it, seems to provide a record of impressive progress in living standards.

The application of modern science to archaeological and anthropological evidence is, however, producing interesting new information on the role of natural resource degradation. Specifically, a pattern of economic and population growth, resource degradation, and subsequent economic decline appears more common than previously thought. A major question of present-day resource management is whether the world as a whole, or some portion of it, might be on a trajectory of this type. A first step in addressing such concerns is to construct a formal model linking population dynamics and renewable resource dynamics.

The primary objective of this paper is to construct such a model. The second objective is to apply this model to the very interesting case of Easter Island which, until recently, has been one of the world's great anthropological mysteries. Our model contains three central components. The first component is Malthusian population dynamics, following Thomas R. Malthus (1798). Malthus argued that increases in real income arising from productivity improvements (or other sources) would tend to cause population growth, leading to erosion and perhaps full dissipation of income gains. He also suggested that population growth might overshoot productivity gains, causing subsequent painful readjustment.[1]

* Brander: Faculty of Commerce, University of British Columbia, 2053 Main Mall, Vancouver, British Columbia, Canada V6T 1Z2; Taylor: Department of Economics, University of British Columbia, 997-1873 East Mall, Vancouver, British Columbia, Canada V6T 1Z1. We thank three anonymous referees for very helpful suggestions. We are also grateful to Bob Allen, Elhanan Helpman, Janis [illegible], Grant McCall, Hugh Neary, Phil Neher, Tobey [illegible], Jim Teller, Guofu Tan, and participants at several conferences and university seminars. Carol McAusland provided valuable research assistance. Financial support from the SSHRC, the Killam Trusts, the Canadian Institute for Advanced Research, and the Network Centre of Excellence for Sustainable Forestry Management is gratefully acknowledged.

[1] Malthus was not a fatalist. He believed that enlightened public policy could reduce population growth, contemplating both contraception and "moral restraint" as

The second component of the model is an open-access renewable resource. Malthus asserted (Ch. 10 pp. 82–83) that the negative effects of population growth would be worse in the absence of established property rights. Property rights are a particular problem with renewable resources such as fish, forests, soil, and wildlife. Thus, if we are to observe Malthusian effects anywhere, we are perhaps likeliest to see them where renewable resources are an important part of the resource base. The extreme version of incomplete property rights is an open-access resource, where anyone can use the resource stock freely. Malthus did not provide a clear formulation of open-access renewable resources, but we can take this step, drawing on the modern theory of renewable resources.

The third component of the model is a Ricardian production structure at each moment in time. Thus the model might reasonably be referred to as a Ricardo-Malthus model of open-access renewable resources. The components of the model are relatively simple. Even so, the model exhibits complex dynamic behavior. For example, one possible outcome of the model is a dynamic pattern in which, starting from some initial state, population and the resource stock rise and fall in damped cycles. A change in parameters, however, can shift dynamic behavior toward monotonic extinction of the population or might lead to monotonic convergence toward an interior steady state.

The model provides a plausible explanation of the rise and fall of the Easter Island civilization. Applying the model to larger and more complex modern resource systems would require an expanded model structure, but the simple model presented in this paper provides, at the very least, insights that should be considered in evaluating current renewable resource management practices.

mitigating factors. He noted that reduced fertility arising from social responses (such as increased age of marriage) was the major check on population in Western Europe. Other checks included reduced fertility and increased mortality arising from poor nutrition and increased incidence of disease. Famine, in his view, was only nature's last resort, and he noted that in much of Europe "absolute famine has never been known" (Malthus, 1798 p. 61).

The main intellectual precursors to our analysis are Malthus (1798), David Ricardo (1817), and the pioneering work on renewable resources of H. Scott Gordon (1954) and M. B. Schaefer (1957). Resource dynamics have been studied by many scholars, and a particularly valuable overview of this material (with considerable original work) is Colin W. Clark (1990). The particular resource model used here is due to Brander and Taylor (1997), and is also related to Anthony D. Scott and Clive Southey (1969). Detailed modeling and estimation of particular renewable resource stocks has been carried out by many scholars including, for example, Jean-Didier Opsomer and Jon M. Conrad (1994). Careful studies of Malthusian population dynamics include Maw-Lin Lee and David Loschky (1987) and George R. Boyer (1989). The claim that environmental constraints are impinging negatively on living standards is a central theme in Richard B. Norgaard (1994) and Lester R. Brown (1995).

Applying formal economic analysis to an archaeological mystery is an unusual activity for economists, but is not without precedent. In particular, Vernon L. Smith (1975) used a formal model of hunting to explain the extinctions of large mammals during the late Pleistocene era and more recently (Smith, 1992) suggested using formal economic models to explain human prehistory more generally. This idea parallels evolution in the field of archaeology itself, where mathematical models are increasingly used as aids to interpreting physical evidence. A valuable overview of this rapidly changing field is Kenneth R. Dark (1995).

Section I provides a brief description of Easter Island's past. This past is fascinating in its own right, as are the methods by which it has been uncovered, but our main goal is to provide background to our approach. Section II presents our general equilibrium model of resource use and Malthusian population dynamics. Section III analyzes population-resource interactions, and states the main propositions characterizing the dynamic behavior of the system. Section IV applies the model to Easter Island and Section V describes other cases where the model might apply. Section VI discusses the role of institutions

change, and Section VII contains concluding remarks.

I. Easter Island

Easter Island (also called Rapa Nui) is a small Pacific island over 2,000 miles (3,200 km) from the coast of Chile, with a population (as of the early 1990's) of about 2,100. For the past two centuries, Easter Island has been regarded as a major archaeological and anthropological mystery. In particular, the Polynesian civilization in place at the time of first European discovery in 1722 is known to have been much poorer and much less populous than it had been a few hundred years earlier. Thus the economic record in Easter Island is one of rising wealth and rising population, followed by decline.

The most visible evidence of Easter Island's past glory consists of enormous statues (called "moai"), carved from volcanic stone. Many statues rested on large platforms at various locations on the island. The largest "movable" statues weigh more than 80 tons, and the largest statue of all lies unfinished in the quarry where it was carved, and weighs about 270 tons. The puzzling feature of the statues and platforms is that the late stone age Polynesian culture found on Easter Island in 1722 seemed incapable of creating such monumental architecture. First, the culture seemed too poor to support a large artisan class devoted to carving statues, and certainly no such group existed in the eighteenth century. Also, the statues were moved substantial distances from the island's stone quarry to their destinations, but the population, estimated at about 3,000 in 1722, seemed too small to move the larger statues,[2] at least without tools such as levers, rollers, rope, and wooden sleds. However, the island in 1722 had no trees suitable for making such tools. Local residents had no knowledge of how to move the statues, and believed that the statues had walked to the platforms under the influence of a spiritual power.

Various theories have been advanced to explain these statues and other aspects of Easter Island. The most well-known theory is due to Thor Heyerdahl (1950, 1989), who argued that native South Americans had inhabited Easter Island (and other Pacific islands), had built Easter Island's statues, and had subsequently been displaced by a less advanced Polynesian culture. To support his thesis, Heyerdahl traveled on a balsa raft, the Kon-Tiki, from off the coast of South America to the Pacific islands. A more exotic theory of the "Atlantis" type, proposed by John Macmillan Brown (1924), is that Easter Island is the tiny remnant of a great continent or archipelago (sometimes called "Mu") that housed an advanced civilization but sunk beneath the ocean. A still more exotic theory, proposed by Erich von Daniken (1970), is that the statues were created by an extraterrestrial civilization. Two recent books describing the current understanding of Easter Island are Bahn and Flenley (1992) and Van Tilburg (1994). This understanding does not support the Heyerdahl, "Atlantis," or extraterrestrial theories of Easter Island, but fits well with the Ricardo-Malthus model of open-access resources.

Recently discovered evidence suggests that Easter Island was first settled by a small group of Polynesians about or shortly after 400 A.D. The pollen record obtained from core samples and dated with carbon dating methods shows that the island supported a great palm forest at this time. This discovery was a major surprise given the treeless nature of the island at the time of first European contact. In the years following initial settlement, one important activity was cutting down trees, making canoes, and catching fish. Thus the archaeological record shows a high density of fish bones during this early period. Wood was also used to make tools and for firewood, and the forest was a nesting place for birds that the islanders also ate. The population grew rapidly and was wealthy in the sense that meeting subsistence requirements would have been relatively easy, leaving ample time to devote to other activities including, as time went on, carving and moving statues.

[2] There is a significant literature devoted to methods of statue construction and movement. Even 3,000 people could not have moved the larger statues without the use of wooden sleds and levers. See Paul Bahn and John Flenley (1992) and Jo Anne Van Tilburg (1994) for a discussion of statue transportation.

Noticeable forest reduction is evident in the pollen record by about 900 A.D. Most of the statues were carved between about 1100 and 1500.[3] By about 1400 the palm forest was entirely gone. Diet changed for the worse as forest depletion became severe, containing less fish (and thus less protein) than earlier. Loss of forest cover also led to reduced water retention in the soil and to soil erosion, causing lower agricultural yields. Population probably peaked at about 10,000 sometime around 1400 A.D., then began to decline.[4] The period 1400 to 1500 was a period of falling food consumption and initially active, but subsequently declining, carving activity.

Carving had apparently ceased by 1500. At about this time, a new tool called a "mataa" enters the archaeological record. This tool resembles a spearhead or dagger and is almost certainly a weapon. In addition, many islanders began inhabiting caves and fortified dwellings. There is also strong evidence of cannibalism at this time. The natural inference is that the island entered a period of violent internecine conflict. However, at first European contact in 1722 no obvious signs of warfare were noted. This visit (by three Dutch ships) lasted only a single day, however, and much may have gone unnoticed.

The next known contact with the outside world was a brief visit from a Spanish ship in 1770, followed in 1774 by a visit from James Cook, who provided a systematic description[5] of Easter Island. There had been some change between 1722 and 1774. Most noticeably, almost all of the statues had been knocked over, whereas many had been standing in 1722. Statue worship, still in place in 1722, had disappeared by 1774. Population was apparently less numerous than it had been in 1722, and was estimated at about 2,000.[6]

Thus Easter Island suffered a sharp decline after perhaps a thousand years of apparent peace and prosperity. The population rose well above its long-run sustainable level and subsequently fell in tandem with disintegration of the existing social order and a rise in violent conflict. Kirch (1984 p. 264) suggests that "Easter Island is a story of a society which—temporarily but brilliantly surpassing its limits—crashed devastatingly."

The mystery of Easter Island's fall is not regarded by many as solved. In simple form the current explanation is that the islanders degraded their environment to the point that it could no longer support the population and culture it once had. However, Polynesians almost always dramatically altered the environments of the islands they discovered. Why did environmental degradation lead to population overshooting and decline on Easter Island but not on the other major islands of Polynesia? Furthermore, there are 12 so-called "mystery islands" in Polynesia. These islands were once settled by Polynesians but were unoccupied at the time of European discovery. All these Polynesian islands represent pieces of "data" that should be consistent with whatever theory is proposed as an explanation for Easter Island.

II. The Ricardo-Malthus Model

A. *Renewable Resource Dynamics*

The resource stock at time t is $S(t)$. For Easter Island it is convenient to think of the resource stock as the ecological complex consisting of the forest and soil. The change in the stock at time t is the natural growth

[3] Radiocarbon dates are available in Van Tilburg (1994 p. 33). These dates are attributed to unpublished work of W. S. Ayres.

[4] This population estimate appears several places in the literature and is often attributed to W. Mulloy. Others have suggested that the reasonable range for the population maximum was between 7,000 and 20,000, with most favoring about 10,000.

[5] Cook had a Tahitian crew member who could communicate quite easily with the Easter Islanders, as Tahitian and the Easter Island language were similar. See P. V. Kirch (1984 Chs. 3 and 11).

[6] In 1862 the population was reliably estimated at about 3,000. In 1862 and 1863, slave traders from Peru invaded the island and took about one-third of the population as slaves. Many of these slaves died of smallpox. A few returned to the island, inadvertently causing a smallpox epidemic that killed most of the remaining Islanders. In [illegible] the population reached its low point of 111, from which it subsequently increased by natural increase and by migration from Tahiti and Chile.

$G(S(t))$, minus the harvest rate, $H(t)$. Dropping the time argument for convenience,

$$dS/dt = G(S) - H. \tag{1}$$

We use the logistic functional form[7] for G, which is perhaps the simplest plausible functional form for biological growth in a constrained environment.

$$G(S) = rS(1 - S/K). \tag{2}$$

K, the "carrying capacity," is the maximum possible size for the resource stock, as $S = K$ implies that further growth cannot occur. Variable r is the "intrinsic" growth (or regeneration) rate.

The economy produces and consumes two goods. H is the harvest of the renewable resource, and M is some aggregate "other good." In the case of Easter Island, we think of the (broadly defined) harvest as being food (i.e., agricultural output from the soil and fish caught from wooden vessels made from trees), while good M would include tools, housing, artistic output (including monumental architecture), household production, etc. Good M is treated as a numeraire whose price is normalized to 1. Aside from resource stock S, the only other factor of production is labor, L. We make the inessential simplification that labor force L is equal to the population. Good M is produced with constant returns to scale using only labor. By choice of units, one unit of labor produces one unit of good M. Since the price of good M is 1, the wage (denoted by w) must equal 1 if manufactures are produced.

We assume that harvesting of the resource is carried out according to the Schaefer harvesting production function (proposed by Schaefer [1957]) as follows,

$$H^P = \alpha S L_H \tag{3}$$

where H^P is the harvest supplied by producers. (The superscript P stands for "production".) L_H is the labor used in resource harvesting and α is a positive constant. Letting $a_{LH}(S)$ represent the unit labor requirement in the resource sector, (3) implies that $a_{LH}(S) = L_H/H^P = 1/(\alpha S)$. Production in both sectors is carried out under conditions of free entry. Because of open access there is no explicit rental cost for using S, and the price of the resource good must equal its unit cost of production.

$$p = w a_{LH} = w/(\alpha S). \tag{4}$$

A representative consumer is endowed with one unit of labor and is assumed to have instantaneous utility given by the following Cobb-Douglas utility function

$$u = h^{\beta} m^{1-\beta} \tag{5}$$

where h and m are individual consumption of the resource good and of manufactures, and β is between 0 and 1. Maximizing utility at a point in time subject to instantaneous budget constraint $ph + m = w$ yields $h = w\beta/p$ and $m = w(1 - \beta)$. Since total domestic demand is L times individual demand, we have

$$H^D = w\beta L/p; \qquad M^D = w(1 - \beta)L \tag{6}$$

where superscript D represents demand.

At a moment in time the resource stock is fixed, the population (and labor force) is fixed, and the economy's production possibility frontier is given by the following full-employment condition.

$$H^P a_{LH}(S) + M = L. \tag{7}$$

A linear production structure of this type is referred to as a Ricardian production structure (after Ricardo, 1817). The Ricardian temporary equilibrium can be determined algebraically by substituting (4) into (6) (i.e., the supply price equals the demand price) to obtain temporary equilibrium resource harvest, H.

$$H = \alpha\beta L S. \tag{8}$$

The equilibrium output of M is $M = (1 - \beta)L$, implying that manufactures will always be produced and therefore that wage $w = 1$. At a Ricardian temporary equilibrium, the

[7] Our major results can be extended to the case of a general compensatory growth function. See Appendix B.

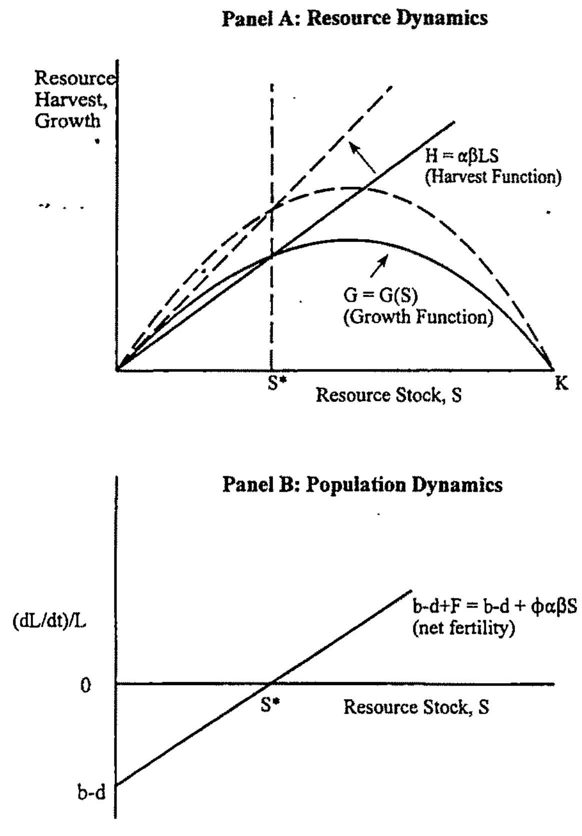


FIGURE 1. A RICARDO-MALTHUS STEADY STATE

harvest will not necessarily equal the underlying biological growth rate of the resource. If, for example, temporary equilibrium harvest H exceeds biological growth G, then the stock diminishes. Substituting (2) and (8) into (1) yields the following expression for the evolution of the resource stock.

$$dS/dt = rS(1 - S/K) - \alpha\beta LS. \tag{9}$$

If the resource stock falls, then labor productivity in the resource sector falls, the Ricardian production possibility frontier shifts in, and this establishes a new temporary equilibrium with a lower harvest.

Panel A of Figure 1 illustrates a typical steady state (i.e., where $dS/dt = 0$) using the harvest function (8) and the resource growth function (2). (Ignore the dashed lines for now.) As shown in the figure, stock S^* implies that harvest $H = \alpha\beta LS^*$ just matches resource growth $G(S)$ at $S = S^*$. Thus S^* is a steady state if the actual population level is L.

B. *Malthusian Population Dynamics*

Our discussion of Figure 1 so far implicitly treats population as fixed at size L, allowing us to focus purely on resource dynamics. We now consider population dynamics. We assume an underlying proportional birth rate, and an underlying proportional death rate. Thus the base rate of population increase is $(b - d)$, which we assume to be negative, implying that without any forest stock or arable soil the population would eventually disappear. However, consumption of the resource good increases fertility and/or decreases mortality, and therefore increases the rate of population growth.[8] In particular, the population growth rate is given by

$$dL/dt = L[b - d + F] \tag{10}$$

where $F = \phi H/L$ is the fertility function, and ϕ is a positive constant. Thus higher per capita consumption of the resource good leads to higher population growth. It is in this sense that population dynamics are "Malthusian." Noting from (8) that $H/L = \alpha\beta S$, equation (10) can be rewritten as

$$dL/dt = L(b - d + \phi\alpha\beta S). \tag{11}$$

III. Population and Resource Interactions

Equations (9) and (11) form a two-equation system of differential equations characterizing the evolution of the Ricardo-Malthus model. These equations are a variation of the Lotka-Volterra predator-prey model.[10] Human population, L, is the "predator" and the resource

[8] One might let net fertility depend on total consumption rather than on resource consumption, although the case could be argued either way and has little effect on the analysis.

[9] Among modern high-income societies, population growth is negatively correlated with income at both the country and individual level. Most premodern societies appear to exhibit Malthusian population dynamics, in which higher consumption causes higher population growth.

[10] Predator-prey systems have sometimes been studied in renewable resource economics. See, in particular, D. L. Ragozin and Gardner Brown (1985) and Philip N. (1990).

stock, S, is the "prey." We concentrate first on steady-state analysis, then turn to the system's dynamic behavior.

A. *Steady-State Analysis*

The system given by (9) and (11) will have a steady state if dL/dt and dS/dt are simultaneously equal to zero. There are three solutions, as expressed in Proposition 1.

PROPOSITION 1: *The Ricardo-Malthus model exhibits three steady states. Steady state 1 is a corner solution at* $(L = 0, S = 0)$. *Steady state 2 is a corner solution at* $(L = 0, S = K)$. *Steady state 3 is an interior solution at*

$$(12) \quad L = [r/(\alpha\beta)] \times [1 - (d - b)/(\phi\alpha\beta K)],$$

$$S = (d - b)/(\phi\alpha\beta).$$

PROOF:

The proof follows immediately from simultaneously setting equations (9) and (11) to zero and can be seen by inspection. To obtain steady state 3, one can solve for S from expression (11), then substitute this value in (9) to obtain L. Note that steady state 3 implies positive values for L and S as $d > b$, and r, α, β, and ϕ are all positive. The steady-state stock, S, must be less than carrying capacity K. In steady state 3, $S = (d - b)/(\phi\alpha\beta)$ so this consistency requirement can be written as

$$(13) \quad (d - b)/(\phi\alpha\beta) < K.$$

If (13) comes to equality or reverses direction, then steady state 3 collapses to steady state 2.

We focus on the interior steady state. As shown in Proposition 4, when (13) holds, the interior steady state is stable whereas steady states 1 and 2 are saddlepoints. The interior steady state can be illustrated graphically by using the two panels of Figure 1. Panel A represents the resource dynamics conditional on population L. Panel B captures the population dynamics given in (11) by graphing the percentage rate of change of the population ($dL/dt)/L$ on the vertical axis, and the resource stock on the horizontal. The upward-sloping line is net fertility ($b - d + F$), which is linear in S as $F = \phi\alpha\beta S$ [substituting (8) into the definition of F].

At any stock $S < S^*$, there is an excess of deaths over births and the population shrinks; at any stock greater than S^* the population grows. At stock S^* population growth is zero and the population level is stationary. We denote this level as L^*. Therefore, at resource stock S^*, the population is stabilized at L^* and the harvest of the resource is just equal to its underlying growth rate, implying that S is also stationary. The two panels together therefore illustrate an interior steady state. Using expression (12), or Figure 1, we can determine how changes in exogenous parameters affect interior steady-state resource stocks and population. Proposition 2 follows by inspection, and Proposition 3 is obtained by differentiating L as given in (12).

PROPOSITION 2: *The steady-state resource stock*

- (*i*) *rises if the mortality rate rises, the birth rate falls, or fertility responsiveness falls;*
- (*ii*) *falls if there is technological progress in harvesting; and*
- (*iii*) *is unaffected by changes in the intrinsic resource regeneration rate, r, or carrying capacity, K.*

PROPOSITION 3: *The steady-state population level*

- (*i*) *rises equiproportionately with an increase in the intrinsic rate of resource growth, r;*
- (*ii*) *falls when harvesting technology improves if* $S < K/2$ *and rises if* $S > K/2$;
- (*iii*) *falls when the taste for the resource good rises if* $S < K/2$ *and rises if* $S > K/2$; *and*
- (*iv*) *rises if the carrying capacity of the environment rises.*

It follows from (8) and (12) that per capita steady-state resource consumption, h, is $(d - b)/\phi$. Per capita consumption of the other

good is $(1 - \beta)$. Thus steady-state resource consumption is determined by demographics, and rises if the birth rate falls or if fertility responsiveness falls.[11] It is instructive to see how population growth dissipates gains from resource productivity improvements. Suppose the economy is at the steady state in Figure 1 and r rises. The growth curve will shift up as shown by the dashed line in Panel A. At stock S^*, the harvest is then less than resource growth and the resource rebuilds. Per capita consumption of the resource good rises, causing population growth, and the harvest function in Panel A rotates upwards (as L rises). Since nothing has altered the demographic steady state in Panel B, the resource stock must return to S^*, but with a higher population and unchanged per capita real income.

B. *Dynamics*

We now characterize the dynamic behaviour of the system. We assume that parameter restriction (13) is met, implying that an interior steady state exists.

PROPOSITION 4: *When an interior steady state exists, the local behavior of the system is as follows.*

(*i*) *Steady state* 1 ($L=0, S=0$) *is an unstable saddlepoint allowing an approach along the* $S = 0$ *axis.*

(*ii*) *Steady state* 2 ($L = 0, S = K$) *is an unstable saddlepoint allowing an approach along the* $L = 0$ *axis.*

(*iii*) *Steady state* 3 ($L > 0, S > 0$) *is a stable steady state and a "spiral node" with cyclical convergence if*

$$(14) \quad r(d-b)/(K\phi\alpha\beta) + 4((d-b) - K\phi\alpha\beta) < 0.$$

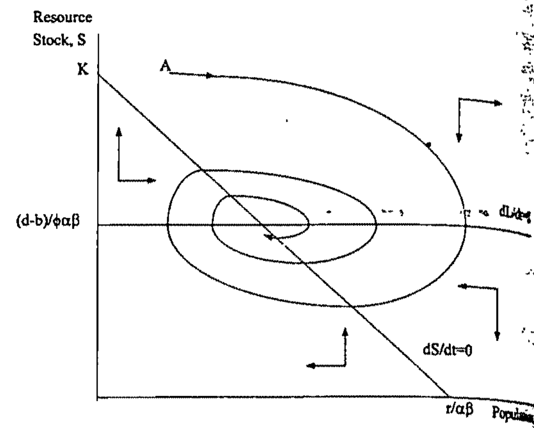


FIGURE 2. TRANSITION DYNAMICS

(*iv*) *Steady state* 3 *is a stable steady state and an "improper node" allowing monotonic convergence if the inequality in* (14) *runs in the other direction.*

PROOF:

See Appendix A.

Proposition 4 is based on the fact that the dynamic system is locally linear in the neighborhood of a steady state. Since Proposition shows that the interior steady state is locally stable, our earlier comparative steady-state exercises (Propositions 2 and 3) are meaningful in that small perturbations in parameters will lead to small changes in steady-state values and allow convergence to a new steady state.

Condition (14) is central in understanding the model's local dynamics. It can be rewritten as $r < 4K\phi\alpha\beta(K\phi\alpha\beta - (d-b))/(d-b)$. Noting from (13) that the right-hand side of this inequality must be positive, one interpretation of (14) is that a slow enough rate of intrinsic resource growth insures a locally cyclical trajectory. Conversely, given r, (14) indicates that cyclical dynamics will occur if fertility is very responsive to per capita consumption (represented by ϕ) or if the harvesting technology is very efficient (i.e., if α is high).

To examine the global properties of the model, consider the phase diagram in Figure 2. Population L is on the horizontal axis. The resource stock S is on the vertical axis. The horizontal line labeled $dL/dt = 0$ derives from

[11] Another easily derived point of interest is that a decline in the birth rate causes steady-state resource output to rise if $S < K/2$. This possibility is unusual for Malthusian models, but arises here because of the resource dynamics.

expression (11), which implies that $dL/dt = 0$ if $S = (d - b)/\phi\alpha\beta$. If the system is above this line, then $dL/dt > 0$, and if the system is below it, then $dL/dt < 0$. The other line, labeled $dS/dt = 0$, is obtained from expression (9), which implies that $dS/dt = 0$ if $S = K - (K\alpha\beta/r)L$. Above this line, $dS/dt < 0$ and below it, $dS/dt > 0$. The intersection of these lines is the interior steady state. The directions of motion for each of the four regions in the diagram are shown by the right-angle arrows.

Consider, for example, point A, which might represent "first arrival" (i.e., a small population and the resource at carrying capacity). Point A is above the $dS/dt = 0$ line, implying that the resource stock must be falling, and is also above the $dL/dt = 0$ line, implying a rising population. The figure shows one possible adjustment path toward the steady state, but other types of adjustment are also consistent with the arrows of motion, including monotonic adjustment toward the steady state. Proposition 5 characterizes the global approach to steady state conditional on different starting points.

PROPOSITION 5: *When an interior steady state exists, the global behavior of the system is as follows.*

(i) If $L > 0$ and $S = 0$, the system approaches steady state 1 with $L = 0$ and $S = 0$.
(ii) If $L = 0$ and $S > 0$, the system approaches steady state 2 with $S = K$ and $L = 0$.
(iii) If $S > 0$ and $L > 0$, then the system converges to the interior solution in steady state 3.

PROOF:

See Appendix A.

It is striking that the system converges to an interior steady state from any interior starting point. There are two important parameter restrictions underlying this property. The first is inequality (13). If this inequality is not satisfied, the system crashes toward zero population. In this case, steady state 2 becomes a globally stable improper node. Thus, if (13) is not satisfied, our model implies extinction of the human population and a restoration of the resource base to carrying capacity.

The second parameter condition is given by (14) which, as shown in Appendix A, determines whether the linearized system in the neighborhood of the interior steady state has complex or real roots. If the linearized system has complex roots, then all trajectories exhibit cyclical adjustment sufficiently close to the steady state. If the linearized system has real roots, then all trajectories approach the interior steady state along a path increasingly close to the dominant eigenvector of the system. Such a trajectory may be globally monotonic and must be locally monotonic.

If we start far away from the steady state, then it is more difficult to describe the paths of the system completely, but many qualitative features of the global system follow from an understanding of the local analysis. For example, suppose the system is perturbed from an initial steady state by the instantaneous disappearance of some fraction of the predator population. Cyclical behavior arises if the predator grows quickly in response to this shock while the resource grows slowly. In this case, the quick growth of the predator causes it to overshoot its new long-run level. The now overabundant predator then reduces the prey below its steady-state level, and this in turn causes a decline of the predator population below its steady-state level. But when the predator declines, the prey rebuilds and overshoots the steady state, leading to a resurgence of the predator, which again overshoots, etc., tracing out a damped cycle with an overshooting predator population chasing a slowly adjusting prey toward the steady state. This interpretation is consistent with condition (14), which shows that adjustment must be cyclical if r (the intrinsic growth rate of the prey) is sufficiently low or if $\phi\alpha\beta$ (the growth response of the predator to a change in the resource stock S) is sufficiently high.

This description describes the forces affecting the local behavior of the system near a steady state, but it also applies to global behavior. A complete analytical characterization of all possible trajectories as a function of parameters is difficult to develop, but the limiting case in which the resource stock adjusts instantaneously to its steady-state

value is instructive. In this case the economy would always operate on the $dS/dt = 0$ locus. As the founding population is (by assumption) smaller than its steady-state value, the economy would move down the $dS/dt = 0$ locus until it reached the steady state. In this case, when the predator adjusts slowly and the prey adjusts infinitely quickly, both the resource stock and the population level adjust monotonically toward their steady-state values.

Technically, we can use the condition $dS/dt = 0$ to solve for S as a function of L [from (9)] then substitute this in (11) to get a one-variable differential equation in L. The solution is logistic: population grows fast at first then levels off, as in the standard description of Polynesian islands.

IV. Applying the Ricardo-Malthus Model to Easter Island

In this section we use simulations of the Ricardo-Malthus model to make two points. First, by choosing parameters that are consistent with our knowledge of Easter Island, Polynesian civilizations generally, and other Neolithic populations, we are able to generate a time series for population size and resource stocks that appears to (approximately) replicate Easter Island's past. We take this as tentative support for our theory of the rise and fall of the Easter Island civilization. Second, parameter changes can change the time pattern of population and resource stock evolution from extreme cyclical overshooting to monotonic adjustment toward the steady state.[12] We take this feature of the model as a possible explanation as to why Easter Island appears to be different from other Polynesian islands.

A. *Parameter Choice*

Some parameter values are simply a matter of scaling, such as the carrying capacity of the forest/soil resource complex. It is convenient for the stock to be similar in magnitude to the population, so we let the carrying capacity of the resource stock be 12,000 units. This is the starting value of the stock when Polynesian colonization first occurred. (The forest had been in place for approximately 37,000 years before first colonization, so carrying capacity had certainly been reached.)

The next parameter to consider is α, labor harvesting productivity. The productivity of a unit of labor is αS. One unit of labor corresponds to the amount of labor one person can provide in one period. It is convenient to let ten-year intervals be periods. If we let $\alpha =$ 0.00001, this means that if $S = K$, a household could provide its subsistence consumption (the amount just necessary to reproduce itself) in about 20 percent of its available labor time. Accordingly, there is considerable surplus on the island when the resource stock is large. This seems roughly consistent with known information.

Parameter β reflects the "taste" for the output of the harvest good. One way of trying to get some idea of β is to recall that β is equal to the share of the labor force devoted to harvesting the resource. The other sector includes manufacturing and service activities. Various pieces of evidence suggest that the resource sector probably absorbed somewhat less than half the available labor supply. A value of 0.4 for β is probably in the reasonable range.

Another important parameter is r, the intrinsic growth (or regeneration) rate of the resource. We initially assume an intrinsic growth rate of 0.04, implying that, left to itself, the forest/soil complex would increase by 4 percent per decade in the absence of congestion effects (i.e., if the stock were small compared to the carrying capacity). The remaining key parameters are the demographic parameters. Let $(b - d) = -0.1$ and let $\phi = 4$. The value for $(b - d)$ means that the population would decrease by 10 percent per decade in the absence of the resource stock. Letting $\phi = 4$ implies that there would be positive population growth if the stock were approximately 50 percent of its carrying capacity, and negative population growth otherwise. Throughout the simulation period annual population growth never exceeds 1 percent per year, which

[12] The key parameter difference that we consider in explaining the difference between Easter Island and other Polynesian islands is substantial. The model also implies bifurcations (in parameter space) around which major changes in predictions arise.

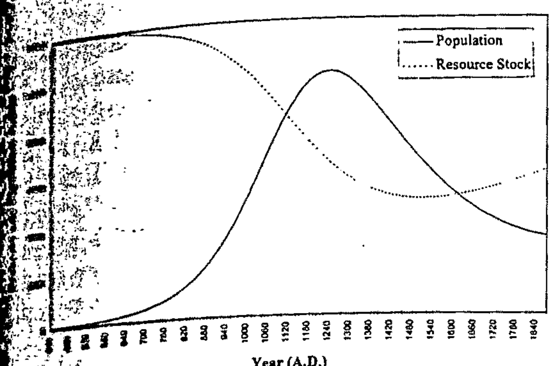


FIGURE 3. EASTER ISLAND BASE CASE

consistent with the demographic literature on Neolithic populations. The range of estimates for the founding population ranges from 20 to 100 or more, with 40 being commonly used as a plausible estimate. We take the starting value to be 40, but varying this estimate makes little difference to the results. These parameters yield the time-series pattern shown in Figure 3.

Figure 3 shows an interesting dynamic pattern. For the first 300 years, humans have little impact on the resource. Population then begins to increase rapidly and the resource stock falls precipitously for the next 800 years. About 800 years after discovery the initial population of 40 has grown to about 10,000. The period of high population (and high labor supply) in the simulation corresponds to the period of intense carving in the archaeological record. The simulated resource stock reaches its trough about 250 years later, close to 1500 A.D., as does per capita resource consumption. Recall that a new weapon (the "mataa") appears in the archaeological record at about this time. Evidence of cannibalism also appears, and there is movement to fortified structures and caves as dwellings. The simulated resource stock begins its recovery but population continues to fall, implying a 1722 population of approximately 3,800 to meet the Dutch ships, not far from the 3,000 actually estimated. The simulated 1774 population is about 3,400, which is somewhat more than the 2,000 Cook estimated. In the 1800's there is substantial outside intervention so our model would no longer hold.

B. *Why Is Easter Island Unusual?*

Our model allows both cyclical and monotonic behavior so it offers the potential to explain both the cyclical overshooting that occurred on Easter Island and the monotonic behavior apparently observed on the major Polynesian islands. There is no reason to believe that Easter Island was an outlier in its underlying demographics, its tastes, or its technology. However, it was an outlier in one very important respect. The palm tree that grew on Easter Island happened to be a very slow-growing palm. This palm (which was the single most significant component of the forest/soil complex on Easter Island) is now known (due to J. Dransfield et al., 1984) to have been a species of Jubea Chilensis (the Chilean Wine palm). This palm tree grows nowhere else in Polynesia, and it is perhaps the only palm that can live in Easter Island's relatively cool climate. An authoritative text (Alexander M. Blombery and Tony Rodd, 1982 p. 110) reports that "Cultivation presents few problems in a suitable temperate climate, but growth of these massive palms is slow and it is generally later generations who get the benefit from their planting." Under ideal conditions, the Jubea palm requires about 40 to 60 years before it reaches the fruit-growing stage, and can take longer.[13]

In contrast, the two most common large palms in Polynesia are the Cocos (coconut palm) and the Pritchardia (Fiji fan palm). Neither of these palms can grow on Easter Island, and both are fast-growing trees that reach fruit-growing age in approximately seven to ten years. For a resource based on these palms, it would be more reasonable that the intrinsic growth rate would be about 0.35 or 35 percent per decade.[14] Figure 4 shows a simulation that

[13] This information is based on private communication with palm growers. Easter Island was also an outlier in rainfall and temperature, contributing to slow growth of the resource.

[14] Translating "time-to-fruit" into intrinsic growth rate r is difficult, as trees continue to grow well after first yielding fruit and we are interested in the entire forest/soil complex in any case. Associating a 40-year time-to-fruit with $r = 0.04$ and a seven–ten-year time-to-fruit with $r = 0.35$ is plausible but very rough. Also, the trees were not

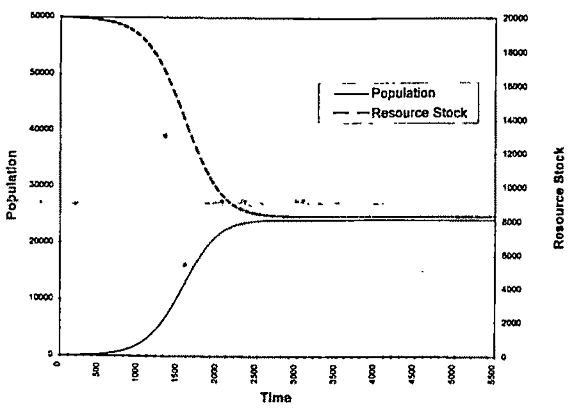


FIGURE 4. POPULATION AND RESOURCE DYNAMICS WITH FAST REGENERATION

is identical to Figure 3 except that the growth rate is raised from 0.04 to 0.35.

The higher intrinsic resource growth rate causes the population to adjust more smoothly. In fact, this simulation is technically cyclical, but the cycle is so muted that the adjustment path is virtually monotonic, as the population peaks at 42,245 before leveling out at its steady-state value of 41,927. The population trajectory would not become strictly monotonic unless the intrinsic growth rate exceeded 0.71, but even at moderate growth rates of 0.15 or 0.2, the population "crash" would be too small to be evident to archeologists. Low growth rates, on the other hand, produce dramatic cyclical fluctuations.

Thus an island with a slow-growing resource base will exhibit overshooting and collapse. An otherwise identical island with a more rapidly growing resource will exhibit a near-monotonic adjustment of population and resource stocks toward steady-state values. Even if everything else were similar across islands, this one fact would allow the Ricardo-Malthus model to be consistent with both the spectacular overshooting and collapse on Easter Island and the far less dramatic development exhibited on other major Polynesian islands.

The model is also consistent with the 12 so-called "mystery islands" that were once settled by Polynesians but were unoccupied at European contact. All but one of these islands have relatively small carrying capacities. Applying our model, we observe that if K is sufficiently small then condition (13) will not be satisfied and there will be no "interior" steady state. A colonizing population could arrive but would eventually drive the resource stock down to a level that would cause extinction of the human population.

Another noteworthy Polynesian settlement is New Zealand's South Island. The South Island had a high concentration of large flightless birds (up to 10 feet tall and 500 pounds in weight) called "Moas." First Polynesian settlement is thought to have occurred around 1000 A.D. (although it may have been later). Following settlement the South Island Maori (or "Moa-hunters") lived "high on the bird" by hunting Moa, along with fishing and agriculture. Over this period there was substantial deforestation and the Moa were driven to extinction. It is not clear exactly when the Moa became extinct, but the larger species disappeared first, possibly lasting as little as 200 years following settlement. There is some disagreement over whether population overshot then declined, or whether it merely stagnated as the Moa disappeared.[15] However, the South Island was more densely settled than the North during the Moa-hunting period, but at European contact (about 1700) settlement was denser in the more temperate and warmer North. Thus it is possible that the South Island exhibited population overshooting, as would be consistent with the slow-growing resource base (consisting of Moa and slow-growing forests).

While Easter Island and Polynesia more generally offer interesting applications of our model, the significance of our analysis would be greatly expanded if the basic approach were also relevant for other cases. In the following section we ask whether population growth and

the only element in the forest/soil complex, which further complicates the problem of estimating the intrinsic growth rate for different hypothetical situations.

[15] Peter Bellwood (1987 p. 157) contends that a population overshoot did occur, stating that "the picture is one of ultimate population decline, with a gradual decrease in cultural energy owing to an insurmountable decrease in resources." A good reference on the Moa is Atholl Anderson (1989).

endogenous resource degradation have played an important role in the decline of other civilizations.

V. Other Applications

Our model might be consistent with the Mayan collapse which, like Easter Island, was long regarded as a mystery. The Mayan empire occupied what is now the Yucatan Peninsula of Mexico; and parts of Guatemala, El Salvador, and Honduras. The empire reached its peak in the period 600–830 A.D., then suffered a rapid decline in both population and cultural sophistication over the following 100 years. The civilization was partially rebuilt in an outlying area but this later Mayan civilization went into decline about 1200 A.D. Only fragments of the Mayan people and civilization still existed at the time of Spanish conquest in 1521.

Recent evidence (including T. Patrick Culbert [1988] and Sarah L. O'Hara et al. [1993]) shows the key role of environmental decline in the Mayan collapse. As on Easter Island, much of the evidence is from carbon-dated core samples showing deforestation, drying, soil erosion and contamination, and reduced crop yields in major agricultural areas. By the early ninth century, agricultural output could no longer support the dense population in the region, leading to out-migration, a sharp decline in population, and a collapse of the sophisticated civilization that had developed. The main unresolved question concerns the extent to which drying and erosion were the endogenous result of human activity rather than arising from exogenous climate changes. The evidence suggests that both factors were important. The following quotes from Culbert (1988 pp. 99–101) provide a now widely accepted interpretation.

> All available data show that populations in the Southern lowlands rose rapidly to a Late Classic peak. Not only was the population unusually dense (200 per sq. km.), but it covered an area too large to allow adjustment through relocation or emigration. ...Maya agriculture became increasingly intensive as population rose.... Overextension—overshoot in systems terminology—often afflicts complex systems during a period of expansion.... In the Maya case, the overextension was ecological and consisted of a population system dependent upon maximal results from a subsistence system that made no allowance for long-term hazards.

Culbert's description of Maya seems similar to the operation of our model. Maya achieved high population density through intensive agriculture that, at best, left no margin of safety when climatic conditions changed marginally for the worse and, at worst, created an endogenously determined agricultural shortfall arising from deforestation and soil erosion.

Resource degradation also played an important role in the decline of the ancient Mesopotamian states (in what is now Iraq). Various civilizations and empires rose and fell in parts of this region between the period 2350 B.C. and 600 A.D. The first true empire was the Akkadian Empire (2350–2150 B.C.). Soil samples from the Akkadian region indicate increasingly intensive agricultural land use until about 2200 B.C., when the soil in the northern part of the empire became too dry to support the population (H. Weiss et al., 1993). The entire northern portion of the empire was abandoned, causing a major migration into the South, which in turn strained food and water supplies in the South to the point of civic collapse and breakdown of central authority (Ann Gibbons, 1993). By 2150 B.C. the empire had degenerated into a group of independent city states.

Later civilizations made extensive use of irrigation. Joseph Tainter (1988) writes "In this area, agricultural intensification and excessive irrigation lead to short-term above-normal harvests, with increasing prosperity ... [but] the rise of saline ground water erodes or destroys agricultural productivity." By the end of the third dynasty of Ur (about 2000 B.C.), agricultural yields per unit of land had fallen by about 50 percent since the first dynasty (about 300 years earlier), and about twice as much seed per unit of land was required even to achieve this lower yield. (See Robert McCormick Adams, 1981 p. 151.) Declining agricultural productivity was an important contributing factor to the decline of Ur.

Other major Mesopotamian civilizations, including the Assyrians, the Babylonians, and

the Sumerians suffered the same salinization problem suffered by the Ur. Gradually, through successive empires based in different parts of the region, nearly the entire area became infertile. According to Tainter (1988), by about 1200 A.D. the total occupied area in the region had fallen to perhaps 5 or 6 percent of its earlier peak, and most previously fertile land was uninhabitable. Furthermore, while climate has fluctuated in the region over the past few thousand years, there is no discernable trend in precipitation (as described by Adams, 1981 pp. 12–13). Therefore, while the relative contribution of natural climate change and human activity to the destruction of a major agricultural region cannot be precisely determined, human agricultural practices seem to have been the dominant factor.

A less well-known example concerns the Chaco Anasazi in the southwestern United States. Between about 1000 A.D. and 1150 A.D. the Chaco Anasazi built an impressive system of roads, settlements, and "great houses." As described in Stephen H. Lekson and Catherine M. Cameron (1995), the largest great houses had about 700 rooms and are thought to have been administrative centers for a complex trading system. In this period, population grew rapidly, probably through immigration as well as natural increase. This region is prone to significant rainfall variations, and a dry period began in 1134, lasting until 1181. After 1134 no new great houses were built, much of the land was abandoned, and the elite culture of the Chaco Anasazi disappeared. The puzzle here is that this drought was no more severe than droughts of the previous century that had been weathered without strain. Tainter (1988) suggests that Chacoan Anasazi economic organization had already been pushed beyond its limits by population growth and that the moderate drought of 1134 pushed an already overloaded system into collapse.

The Maya, the Anasazi, and the ancient Mesopotamian civilizations all show a similar pattern. In each case, decline of the resource base, particularly soil degradation, was the main factor precipitating a population crash and the decline of a complex civilization. While exogenous climate fluctuations may have played a significant role in these cases, population growth and endogenous resource degradation were also important, making them similar to Easter Island.[16]

Overall, evidence on soil change and core information recently obtained from core samples has significantly changed the way modern archaeologists interpret the past. Resource degradation is now understood to be common in major civilizations. The role of warfare and violent conflict is also being reinterpreted. Rather than being the cause of decline, violent conflict is commonly the result of resource degradation and occurs after the civilization has started to decline, as on Easter Island.

VI. Institutional Adaptation

A critic might object that our analysis underestimates institutional adaptation. We assume an open-access resource, but perhaps we should expect more efficient resource management institutions to evolve. This is primarily an empirical question.[17] Elinor Ostrom (1990) has studied the historical record on common property problems and argues persuasively that efficient institutional reforms sometimes occur in primitive (and advanced) societies, but sometimes do not. (See also Ostrom et al., 1994.) In Ostrom (1990 p. 21) she writes "some individuals have broken out of the trap inherent in the commons dilemma, whereas others continue remorsefully trapped into destroying their own resources." She also observes (1990 p. 210) that "we cannot [adopt] ... a presumption that appropriators will adopt new rules whenever the net benefits of a rule change will exceed net costs."

The main objective of Ostrom (1990) is to determine the factors that favor efficient institutions and those that impede efficient

[16] It is also possible that exogenous climate changes may have affected Easter Island, as emphasized to us in private communication by Grant McCall. See also J. Flenley et al. (1991).

[17] Various theoretical arguments can also be advanced to explain why a society may not undertake an efficiency-enhancing policy change. Raquel Fernandez and Dani Rodrik (1991) model a status quo bias against such reforms that arises when individual gainers and losers from reform cannot be readily identified *ex ante*. Similarly, Alberto Alesina and Allan Drazen (1991) provide a model in which reforms are delayed if different groups can attempt to shift the burden of adjustment to other groups.

tional response. The most important favorable factor is an agreed-upon and correct understanding of the problem. If a soil erosion problem is falsely attributed to low rainfall, then the response might be more rain dances rather than restructured property rights. To use a modern example, it was not possible to settle on an effective response to ozone depletion until there was substantial agreement that a problem existed and on the mechanism causing the problem. Even then, obtaining consensus was difficult, and several major countries have refused to participate in the resulting international agreement.

It is also helpful if proposed rule changes affect relevant parties in a similar way rather than generating winners and losers. Other favorable factors include low discount rates and low enforcement costs. It is also helpful if the affected group is small and if the group has a high level of initial trust and sense of community. Thus, for example, if there are existing ethnic or social divisions that dominate the way people perceive issues, this makes appropriate institutional change difficult to achieve. These conditions follow from the general principle that institutional change is more likely to occur when the individuals who must make the change are confident that they will be among the beneficiaries.

Easter Island did not present a favorable environment for efficient institutional change. It is unlikely that the Islanders understood the biology of the forest-soil complex or the likely incentive effects of alternative institutional arrangements. It is even possible that individual Islanders did not recognize that depletion was taking place. Although the forest disappeared rapidly by archaeological standards, change was slow over the course of an individual life span. Typical life spans for those who survived infancy would have been on the order of 30 years, and even during most rapid depletion, the forest stock would have declined by no more than 5 percent over a typical lifetime. Even if the problem had been recognized, the 40 to 60 years taken for a tree to mature would exceed the working life of virtually all islanders. Thus a program of replanting and caring for seedlings would almost never have been of direct benefit to the cultivators.

The Easter Islanders did make some institutional changes in response to resource scarcity. One major change was that at some point the Easter Islanders abandoned the system of statue worship and apparently pushed over almost all of the statues (usually facedown). Institutional changes of this type, amounting to religious revolution, are clearly very costly and, while they are understandable responses to declining circumstances, they are unlikely to have helped the underlying resource-use problem.

The other kind of institutional reform suggested by our model is population control. There is a large and fascinating literature on social institutions affecting fertility. Perhaps the main point is simply that the range of such institutions is very large. Some societies adopted practices that directly limited population growth, including infanticide and genital mutilation. More subtle and benign approaches involving marriage customs and crude contraceptive methods were also used. Such population controls tend to be density dependent. For example, as crowding increases and resources become more scarce, it is more difficult for young men to acquire sufficient wealth to marry, and both infanticide and contraception would be practiced more frequently. This was the pattern elsewhere in Polynesia and was probably true of Easter Island as well. This is consistent with our model, as it means that net fertility falls when per capita resource consumption falls.

However, in societies that lack a clear scientific understanding of their world, institutional adaptation involving fertility, property rights, and other matters is likely to be a "trial and error" process. Efficient institutions would probably be achieved only after a long period of time and many trials. It would be difficult for a society like Easter Island to adapt efficiently in a single boom and bust cycle.

These longevity estimates are derived from the four major skeletal series available for prehistoric Polynesia. We have used the life tables in Kirch (1984 pp. 112–14) to calculate that life expectancy at age 5 is slightly less than 30 years, and that the percentage surviving until their mid-forties would be approximately 5 percent.

The other declining civilizations also seem to be poor candidates for efficient institutional reform. We do not have space to discuss each case in detail, but anticipation and understanding of the ecological problem would have been limited, and conflict between competing groups would have delayed reform. Even in sophisticated societies like Maya and Mesopotamia, where property rights may have been relatively secure, the complex array of externalities involved in soil erosion, salinity and declining water tables would have been hard to address successfully.

We recognize, however, that modern knowledge of institutions in preliterate civilizations is very limited. In considering Polynesia, we know there was substantial variation across the different islands. Therefore, one alternative hypothesis to ours is that institutional variation across societies within Polynesia might be the explanation for contrasting growth experiences.

VII. Concluding Remarks

This paper presents a simple model of renewable resource growth and population dynamics and employs this model to provide a plausible account of the rise and fall of the Easter Island civilization. For reasonable parameter values, the model generates a boom and bust cycle in which population grows, the resource base is degraded, and population ultimately falls. This cycle arises because the resource base has a slow regeneration rate. A faster-growing resource would allow monotonic convergence toward the steady-state population. Thus the model can explain the difference between Easter Island and other Polynesian islands based on known differences in resource growth rates.

Our analysis has several lessons for the modern world. First, the model implies that changes in technology, the environment, or human behavior can create feast and famine cycles that may be a recipe for violent conflict over apparently diminishing resources. Easter Island may be only one case of many where unregulated resource use and Malthusian forces led to depletion of the resource base and social conflict. Identifying countries at risk in the modern world may be difficult, but our model provides some guidance as it identifies the key parameters that make cyclical downturns more likely.

A modern case that might be consistent with our model is Rwanda, which entered the news during 1994 because of a violent civil war. This war was normally attributed to ethnic tensions between Hutus and Tutsis, but more careful analysis suggests the possibility that Malthusian population growth, resource degradation, and resulting competition for resources was at the root of the conflict. Between 1950 and 1994, population in Rwanda quadrupled. The boom began in the 1950's when advances in health care and agricultural practice led to increasing real incomes and rising net fertility. By the 1980's what had been an open frontier was "filled up," and real living standards started to fall. Conflict over land between Hutus and Tutsis became increasingly severe, culminating in a civil war in which a significant fraction of the population was killed and a very large fraction (perhaps [illegible] percent) became refugees. Thomas Homer-Dixon (1994) describes several other modern cases where resource degradation driven by population growth has caused violent conflict and local decline of living standards. Models of the Easter Island type, with explicit resource and demographic dynamics, might be helpful in understanding such situations.

A second lesson of our analysis is that it provides one of the first formal empirical examples in which the cycles that arise in nonlinear models appear relevant in analyzing long-run economic development. In short, it suggests that nonlinear dynamics are likely to be relevant in studying economic growth, especially in situations where renewable resources are important.

Third, our analysis of Easter Island and the other cases suggests that economic decline based on natural resource degradation is not uncommon. Institutional change could potentially have averted collapse in many of these societies but it was not undertaken (or at least was not undertaken fast enough). Institutional failure in renewable resource use does happen and it has been fatal for several societies. Recent events in the world's major fisheries suggest that institutional change remains difficult. An extreme case is the (Canadian) Newfoundland cod fishery which was closed in 1992,

cod stocks were down to less than 5 percent of their 1960 levels. Our work offers support for the position that it is both important and difficult to reach efficient institutional arrangements in renewable resource use.

Finally, despite our model's rather gloomy implications, we do not wish to embrace the pessimism of modern neo-Malthusians. First, in considering the modern world, one would introduce nonlinearity in the response of fertility to consumption so as to allow for a demographic transition of the type observed in modern high-income societies. Specifically, net fertility declines with income at sufficiently high-income levels. In our model, cyclical dynamics arise only when fertility has a strong enough positive response to per capita consumption, so incorporating a demographic transition would probably allow the possibility of escape from cyclical dynamics if a high enough income could be reached.

In addition, our model abstracts from technological progress, which is the main force emphasized by growth optimists. Abstracting from technological progress is reasonable for our discussion of Easter Island, and perhaps for most of the other examples we have discussed, but it would clearly be a serious omission in considering the modern world. The model can, however, be readily augmented in this direction. Both the r parameter (resource growth) and the α parameter (harvesting efficiency) could be viewed as susceptible to either exogenous or endogenous technical progress. Furthermore, progress in the form of further scientific understanding may also facilitate institutional adaption in resource use. This would require a larger modification in the model as it implies a different characterization of temporary equilibrium in the resource sector. On the other hand, the possibility of demographic transition notwithstanding, net fertility (particularly declining mortality) is affected by technical progress, and increases in net fertility driven by improvements in medical technology may well lead to population overshooting.

Finally, while technical progress has been the dominant force in the growth process since the beginning of the industrial revolution, this last 200 years is a small fraction of the time that humans have harvested from the earth and built complex, but ultimately fragile, societies. It is, for that matter, shorter than what might be regarded as the "golden age" of Easter Island.

APPENDIX A

PROOF OF PROPOSITION 4:

Let (L^*, S^*) represent a steady state. Define the vector $\mathbf{u} = (u_L, u_S) = (L - L^*, S - S^*)$. Thus $\mathbf{u}$ is the vector of deviations in L and S from a particular steady state. It follows that $du_L/dt = dL/dt$ and $du_S/dt = dS/dt$, where dL/dt and dS/dt are given by (11) and (9). Using a Taylor series expansion for du/dt around $\mathbf{u} = 0$ [i.e., around (L^*, S^*)], it can be shown (as in William E. Boyce and Richard C. DiPrima, 1992 pp. 450–51) that du/dt can be expressed as follows.

$$\text{(A1)} \qquad du/dt = \mathbf{J}(L^*, S^*)\mathbf{u} + R(L, S)$$

where $\mathbf{J}$ is the Jacobian matrix of first-order partial derivatives of dL/dt and dS/dt with respect to L and S, and $R(L, S)$ is a remainder of higher-order terms that can be ignored near $\mathbf{u} = 0$. $\mathbf{J}$ is evaluated at (L^*, S^*). Denoting the components of $\mathbf{J}$ as J_{11}, J_{12}, etc., in the obvious way, we can write this linear system as

$$\text{(A2)} \qquad du/dt = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} \begin{bmatrix} u_L \\ u_S \end{bmatrix}.$$

A two-equation system of linear differential equations [as in (A2)] has a general solution of the form

$$\text{(A3)} \qquad \mathbf{u}(t) = c_1 E_1 e^{z1t} + c_2 E_2 e^{z2t}$$

where c_1 and c_2 are constants, $z1$ and $z2$ are the eigenvalues of coefficient matrix $\mathbf{J}$, and $E1$ and $E2$ are the corresponding eigenvectors. The dynamic behavior of the system depends on whether $z1$ and $z2$ are real, complex, or imaginary, and on whether any real part of these eigenvalues is positive or negative. The system is explosive if $z1$ and $z2$ are positive real numbers, and converges monotonically to the steady state if $z1$ and $z2$ are negative. If $z1$ and $z2$ are complex numbers then cyclical behavior emerges. The coefficients of $\mathbf{J}$ can be

determined by taking partial derivatives of (9) and (11).

(A4) $J_{11} = (b-d) + \phi\alpha\beta S$; $J_{12} = \phi\alpha\beta L$;

$J_{21} = -\alpha\beta S$; $J_{22} = r - 2rS/K - \alpha\beta L$.

(i) For steady state 1 ($L = 0$, $S = 0$), coefficient matrix **J** becomes

$$\text{(A5)} \qquad \mathbf{J}(0,0) = \begin{bmatrix} b-d & 0 \\ 0 & r \end{bmatrix}.$$

The eigenvalues are the diagonal elements $b - d < 0$ and $r > 0$. This combination of one negative real eigenvalue and one positive real eigenvalue implies that steady state (0, 0) is an unstable saddlepoint.

(ii) and (iii) For steady state 2 ($L = 0$, $S = K$) and steady state 3 ($L = (r/\alpha\beta)(1 - S/K)$, $S = (d - b)/\phi\alpha\beta)$) we proceed in the same way, making the appropriate substitutions in matrix **J** using (A4) and calculating the associated eigenvalues of **J**. For steady state 2 the eigenvalues are $-r$ and $(b - d) + \phi\alpha\beta K$ [which is positive by (13)]. As with steady state 1, the combination of positive and negative real eigenvalues implies that steady state 2 is an unstable saddlepoint. For steady state 3, matrix **J** has nonzero off-diagonal elements, so the eigenvalues cannot be seen by inspection but must be obtained as the roots of the characteristic equation. Letting S^* denote the steady-state value of S, the characteristic equation is $z(rS^*/K + z) - (d - b)r(1 - S^*/K) = 0$, which is quadratic in z and has roots

$$\text{(A6)} \qquad z = [-rS^*/K \pm ((rS^*/K)^2 - 4(d-b)r(1 - S^*/K))^{1/2}]/2.$$

If the discriminant of (A6) is positive (or zero), then both solutions for z must be negative real numbers and the steady state is a stable node with monotonic convergence. A negative discriminant implies complex eigenvalues with a negative real part, and the steady state is a stable spiral point. The system then exhibits damped cycles that converge on the steady state. Noting that equation (14) given in the proposition is simply the discriminant of characteristic equation (A6), the proof is complete.

PROOF OF PROPOSITION 5:

(i) Proposition 4 establishes that steady state 1 ($L = 0$, $S = 0$) is a saddlepoint allowing a local approach along the horizontal axis, and we can tell from inspection of the full nonlinear system that steady state 1 is reached from any point along the horizontal axis. If $S = 0$, and $L > 0$, then L must fall to zero. Thus one trajectory of our full nonlinear system is given by the horizontal axis in Figure 2.

(ii) We know from Proposition 4 that steady state 2 ($L = 0$ and $S = K$) is a saddlepoint allowing a local approach along the vertical axis, and we can tell from inspection that steady state 2 is reached from any point along the vertical axis. That is, if $L = 0$, and $S > 0$, then S must raise toward level K, hence another trajectory of our full nonlinear system is given by the vertical axis in Figure 2.

(iii) As the differential equation system is autonomous and continuously differentiable, no two trajectories of the system can intersect. Any trajectory that starts from a point strictly interior in Figure 2 must remain strictly interior. (Otherwise it would touch one of the axes, which we know is impossible, as the axes themselves are trajectories.) We can then divide equation (9) by (11) to obtain the slope of any system trajectory as $dS/dL = [rS(1 - S/K) - \alpha\beta LS]/[L(b - d + \phi\alpha\beta S)]$. Inspection of this slope in each region of the phase diagram in Figure 2 implies that the direction of any trajectory must eventually be inward towards the interior steady state. Limit cycles can be ruled out by applying a theorem due to Kolmogorov as provided by Robert M. May (1973 pp. 85–89). Therefore, the trajectory must approach the steady state.

APPENDIX B

The paper uses the logistic growth function, but the analysis is readily generaliz

to a general compensatory (bent over) growth function provided that $G(0) = G(K) = 0$ and that $G(S)/S$ is strictly decreasing in S. In this case we let $r = \lim_{S \to 0} G(S)/S$. Equation (9) becomes $dS/dt = G(S) - \alpha\beta LS$. Equation (11) is unchanged. Proposition 1 follows immediately, except that in steady state 3, the steady-state value of L is expressed as $L^* = G(S^*)/(S^*\alpha\beta)$, where $S^* = (d-b)/(\alpha\beta\phi)$ as before. Proposition 2 follows immediately. To obtain an analog of Proposition 3, we need to replace the experiment of increasing r by an overall increase in $G(S)/S$ and, instead of focussing on whether $S < K/2$ or $S > K/2$, we focus on whether $G(S)$ is increasing or decreasing. Figures 1 and 2 have the same general form as before. In Proposition 4, parts (i) and (ii) are unchanged, and (iii) follows as before except that condition (14) becomes $H(S^*)^2 - 4(d-b)G(S^*)/S^* \leqslant 0$, where $H(S^*) = G'(S^*) - G(S^*)/S^*$. Proposition 5 is unchanged. To carry out the Easter Island simulation we, of course, require some specific functional form. The logistic form works well, but it is clear from this brief discussion that other forms would also work. More general analysis of general functional structures is difficult, although methods along the lines of those used in Peter Howitt and R. Preston McAfee (1988) could perhaps be applied.

REFERENCES

Adams, Robert McCormick. *Heartland of cities: Surveys of ancient settlement and land use on the central floodplain of the Euphrates*. Chicago: University of Chicago Press, 1981.

Alesina, Alberto and Drazen, Allan. "Why Are Stabilizations Delayed?" *American Economic Review*, December 1991, *81*(5), pp. 1170–88.

Anderson, Atholl. *Prodigious birds: Moas and moa-hunting in prehistoric New Zealand*. Cambridge: Cambridge University Press, 1989.

Bahn, Paul and Flenley, John. *Easter Island earth island*. London: Thames and Hudson, 1992.

Bellwood, Peter. *The Polynesians: Prehistory of an island people*. London: Thames and Hudson, 1987.

Blombery, Alexander M. and Rodd, Tony. *Palms: An informative practical guide to palms of the world, their cultivation, care and landscape use*. London: Angus and Robertson, 1982.

Boyce, William E. and DiPrima, Richard C. *Elementary differential equations and boundary value problems*, 5th Ed. New York: Wiley, 1992.

Boyer, George R. "Malthus Was Right after All: Poor Relief and Birth Rates in Southeastern England." *Journal of Political Economy*, February 1989, *97*(1), pp. 93–114.

Brander, James A. and Taylor, M. Scott. "International Trade and Open Access Renewable Resources: the Small Open Economy Case." *Canadian Journal of Economics*, August 1977, *30*(3), pp. 526–52.

Brown, John Macmillan. *The riddle of the Pacific*. London: Fisher Unwin, 1924.

Brown, Lester R. "Nature's Limits," in Lester R. Brown, ed., *State of the world 1995*. New York: Norton, 1995, pp. 3–20.

Clark, Colin W. *Mathematical bioeconomics: The optimal management of renewable resources*, 2nd. Ed. New York: Wiley, 1990.

Culbert, T. Patrick. *The collapse of ancient states and civilizations*. Tucson, AZ: University of Arizona Press, 1988.

Daniken, Erich von. *Chariots of the gods? Unsolved mysteries of the past*. New York: Putnam, 1970.

Dark, Kenneth R. *Theoretical archaeology*. Ithaca, NY: Cornell University Press, 1995.

Dransfield, J.; Flenley, J. R.; King, S. M.; Harkness, D. D. and Rapu, S. "A Recently Extinct Palm from Easter Island." *Nature*, December 1984, *312*(4986), pp. 750–52.

Fernandez, Raquel and Rodrik, Dani. "Resistance to Reform: Status Quo Bias in the Presence of Individual-Specific Uncertainty." *American Economic Review*, December 1991, *81*(5), pp. 1146–55.

Flenley, J.; King, S.; Teller, J.; Prentice, M.; Jackson, J. and Chew, C. "The Late Quaternary Vegetational and Climatic History of Easter Island." *Journal of Quaternary Science*, 1991, *6*(1), pp. 8–115.

Gibbons, Ann. "How the Akkadian Empire Was Hung Out to Dry." *Science*, August 1993, *261*(5122), p. 985.

Gordon, H. Scott. "The Economic Theory of a Common-Property Resource: The Fishery." *Journal of Political Economy*, April 1954, *62*(2), pp. 124–42.

Heyerdahl, Thor. *The Kon-Tiki expedition*. London: Allen and Unwin, 1950.

———. *Easter Island: The mystery solved.* London: Souvenir Press, 1989.

Homer-Dixon, Thomas. "Environmental Scarcities and Violent Conflict: Evidence from Cases." *International Security*, Summer 1994, *19*(1), pp. 5–40.

Howitt, Peter and McAfee, R. Preston. "Stability of Equilibria with Externalities." *Quarterly Journal of Economics*. May 1988, *103*(2), pp. 261–77.

Kirch, P. V. *The evolution of the Polynesian chiefdoms*. Cambridge: Cambridge University Press, 1984.

Lekson, Stephen H. and Cameron, Catherine M. "The Abandonment of Chaco Canyon, the Mesa Verde Migrations, and the Reorganization of the Pueblo World." *Journal of Anthropological Archaeology*, June 1995, *14*(2), pp. 184–202.

Lee, Maw-Lin and Loschky, David J. "Malthusian Population Oscillations." *Economic Journal*, September 1987, *97*(387), pp. 727–39.

Malthus, Thomas R. *An essay on the theory of population*. Oxford: Oxford University Press, 1798.

May, Robert M. *Stability and complexity in model ecosystems*. Princeton, NJ: Princeton University Press, 1973.

Neher, Philip. *Natural resource economics: Conservation and exploitation*. Cambridge: Cambridge University Press, 1990.

Norgaard, Richard B. *Development betrayed: The end of progress and a coevolutionary revisioning of the future*. New York: Routledge, 1994.

O'Hara, Sarah L.; Street-Perrott, F. Alayne and Burt, Timothy P. "Accelerated Soil Erosion Around a Mexican Highland Lake Caused by Prehispanic Agriculture." *Nature*, March 1993, *362*(6415), pp. 48–50.

Opsomer, Jean-Didier and Conrad, Jon M. "An Open-Access Analysis of the Northern Anchovy Fishery." *Journal of Environmental Economics and Management*, July 1994, *27*(1), pp. 21–37.

Ostrom, Elinor. *Governing the commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press, 1990.

Ostrom, Elinor; Gardner, Roy and Walker, James. *Rules, games, and common-pool resources*. Ann Arbor, MI: University of Michigan Press, 1994.

Ricardo, David. *Principles of political economy and taxation*. (Reprinted) London: Dent, 1817.

Ragozin, David L. and Brown, Gardner. "Harvest Policies and Nonmarket Valuation in a Predator-Prey System." *Journal of Environmental Economics and Management*, June 1985, *12*(2), pp. 155–68.

Schaefer, M. B. "Some Considerations of Population Dynamics and Economics in Relation to the Management of Marine Fisheries." *Journal of the Fisheries Research Board of Canada*, 1957, *14*, pp. 669–81.

Scott, Anthony D. and Southey, Clive. "The Problem of Achieving Efficient Regulation of a Fishery," in Anthony D. Scott, ed., *The economics of fishery management: A symposium*. Vancouver, Canada: Institute of Animal Resource Ecology, University of British Columbia, 1969, pp. 47–59.

Smith, Vernon L. "The Primitive Hunter Culture, Pleistocene Extinction, and the Rise of Agriculture." *Journal of Political Economy*, August 1975, *83*(4), pp. 727–55.

———. "Economic Principles in the Emergence of Humankind: Presidential Address to the Western Economic Association." *Economic Inquiry*, January 1992, *30*(1), pp. 1–13.

Tainter, Joseph. *The collapse of complex societies*. Cambridge: Cambridge University Press, 1988.

Van Tilberg, Jo Anne. *Easter Island: Archaeology, ecology, and culture*. London: British Museum Press, 1994.

Weiss, H.; Courty, M. A.; Wetterstrom, W.; Guichard, F.; Senior, L.; Meadow, R. and Curnow, A. "The Genesis and Collapse of Third Millennium North Mesopotamian Civilization." *Science*, August 1993, *261*(5122), pp. 995–1004.

Sources of Inefficiency in a Representative Democracy: A Dynamic Analysis

By TIMOTHY BESLEY AND STEPHEN COATE *

This paper studies the efficiency of policy choice in representative democracies. It extends the citizen-candidate model of democratic policy-making to a dynamic environment. Equilibrium policy choices are shown to be efficient in the sense that in each period, conditional on future policies being selected through the democratic process, there exists no alternative current policy choices which can raise the expected utilities of all citizens. However, policies that would be declared efficient by standard economic criteria are not necessarily adopted in political equilibrium. The paper argues that these divergencies are legitimately viewed as "political failures." (*JEL* D61, D78, H11)

One of the crowning achievements of neoclassical economics is a rigorous appreciation of the performance of markets in allocating resources. However, for resources allocated in the public sector, our understanding is much less complete.[1] In part this reflects the lack of a satisfactory theoretical framework to analyze policy choice in representative democracies. In Besley and Coate (1997), we introduced a model of democratic policy making as a candidate framework for the systematic analysis of these questions. In this model, policy authority is delegated to particular citizens and individuals compete, through the electoral process, to acquire this power.[2]

This paper uses the model to investigate the efficiency of equilibrium policy choices with repeated elections. Efficiency issues are then more subtle because preferences extend over the entire future policy sequence, while policy makers can control only what happens in their current term. We find that, while political equilibrium does satisfy a certain efficiency property, this does not imply that policies are efficient according to standard economic criteria. We also discuss whether the nonimplementation of economically efficient policies can legitimately be viewed as "political failure," drawing the parallel with the market failure literature.

To analyze policy choices in a dynamic setting, we embed our model of representative democracy in a simple two-period economic model that incorporates redistribution and public investment. A single consumption good is produced using labor and citizens obtain utility from consumption and leisure, differing in their productive abilities. Redistributive taxation takes the form of a linear income tax as in Thomas Romer (1975) and Alan H. Meltzer

* Besley: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE; Coate: Department of Economics, University of Pennsylvania, Philadelphia, PA 19104. The authors are grateful for comments from two anonymous referees, Avinash Dixit, Tim Groseclose, Kenneth Koford, John Lott, George Mailath, Gian Maria Milesi-Ferretti, and Stephen Morris, as well as numerous seminar participants. Besley is also grateful for the hospitality of the Research School of Social Sciences, Australian National University, where his work on the first draft of this paper was completed.

[1] Despite the vast volume of research on public choice, there is no clear consensus on the ability of representative democracy to produce efficient outcomes. At one extreme are writers in the Chicago tradition, such as George Stigler (1982), Gary Becker (1985), and Donald Wittman (1989), who argue that political competition will bring about efficient policy choices. At the other extreme are James M. Buchanan, Gordon Tullock, and their followers (the "Virginia School"), who see "political failures" as pervasive.

[2] A similar model, which shares the basic idea of viewing candidates for political office as citizens, was introduced independently by Martin J. Osborne and Al Slivinski (1996). They coined the term "citizen-candidates" to describe the approach.

and Scott F. Richard (1981). The public investment, which can be thought of as infrastructure or education, enhances citizens' second-period productivities.[3] In each period, the citizen selected to be policy maker chooses the parameters of the tax system. The first-period policy maker must also decide whether to implement the public investment.

We show that equilibrium policy choices are efficient in the following sense; in each period, conditional on future policies being selected through the democratic process, there exist no alternative current policy choices that can raise the expected utilities of all citizens. To show that this does not imply that equilibrium policy choices are efficient according to standard economic definitions, we focus on the public investment decision. We identify three reasons why public investments that are potentially Pareto improving with the available policy instruments may not be undertaken in political equilibrium. Each stems from the problem that when a society makes policy decisions via representative democracy it cannot commit to future policy outcomes.

The first reason concerns nonpayment of future compensation. To generate a Pareto improvement, a public investment may require some individuals to be compensated, via the tax and transfer system, after future gains are realized. However, compensation may not actually be paid in political equilibrium, since future tax rates are determined by policy makers at that time. Thus, if future policy makers are expected to have different preferences, the incumbent may be deterred from undertaking efficient investments.

The other two reasons apply to public investments that do not require manipulation of future taxes and transfers to be Pareto improving. These may not be undertaken if they change the *identity* of future policy makers in a way disadvantageous to the current incumbent. This can happen when an investment, by altering citizens' productive abilities, leads to changes in preferences for redistribution, resulting in the election of a citizen with different policy preferences. Public investments that change the *choices* of future policy makers may also not be undertaken. By changing the income distribution, a public investment can alter the desired tax rates of future policy makers. Such changes can deter an incumbent from undertaking an investment.

To assess whether the nonimplementation of potentially Pareto-improving investments is legitimately viewed as a "political failure" requires us to take a stance on what constitutes such a failure. In the interests of creating a level playing field for markets and governments, a sensible definition of political failure should parallel that used in a market context. In this sense, we argue that failing to undertake a public investment that is potentially Pareto improving with the available policy instruments does constitute a political failure.[4]

Our analysis is related to a number of papers in the macro-policy literature. Torsten Persson and Lars Svensson (1989) argue that debt policy can stray from the efficient path in political equilibrium. Current incumbents run budget deficits to manipulate the choices of future policy makers who do not share their policy preferences. This is one reason for political failure in our model.[5] Philippe Aghion and Patrick Bolton (1990) and Gian Maria Milesi-Ferretti and Enrico Spolaore (1994) develop models in which policy is distorted because current policy choices affect which political party will win in the future. This incentive is also present in our model.[6]

[3] For reasons that are subsequently discussed, the problems we identify are less likely to apply to public investments that produce future consumption benefits.

[4] This paper builds on the analysis of inefficiency in a static model in Besley and Coate (1997), which shows how political competition could fail to secure the election of the most competent policy makers. Imperfect information can also generate inefficiencies as in Kenneth Rogoff (1990) and Coate and Stephen Morris (1995).

[5] In a related argument, Guido Tabellini and Alberto Alesina (1990) show that, in an environment of political instability, those currently holding political power have an incentive to borrow from the future (i.e., run deficits) because they can control how such resources are allocated. See also Amihai Glazer (1989).

[6] Most existing models rest on incomplete approaches to the electoral process. Persson and Svensson (1989) offer no political model; they assume that current policy makers anticipate future policy makers with different preferences. Aghion and Bolton (1990) and Milesi-Ferretti and Spolaore (1994) use a model of two parties with

The organization of the remainder of the paper is as follows. The model is presented in Section I. In Section II political equilibrium is defined and equilibrium policy choices are characterized. Section III analyzes the efficiency of the policy choices emerging from political equilibrium, focusing largely on the investment decision. Section IV discusses the interpretation of the results, while Section V concludes.

I. The Model

A. *The Economic Environment*

The economy consists of N citizens, indexed by $i \in \mathcal{N} = \{1, \dots, N\}$, and lasts for two periods, indexed by $\tau \in \{1, 2\}$. There is a single (nonstorable) consumption good, denoted by x, that is produced using a single factor, labor, denoted by ℓ. Each citizen is endowed with one unit of labor in each period. Citizens differ in their productive abilities. If citizen i supplies $\ell_{\tau i}$ units of labor in period τ, he produces $x_{\tau i} = a_{\tau i}\ell_{\tau i}$, where $a_{\tau i}$ is his period τ "ability." The consumption good is produced competitively with wage rates being equal to abilities for all citizens. Citizen i's period τ utility is $u(x_{\tau i}, \ell_{\tau i})$, where the common per period utility function $u(\cdot)$ is smooth, strictly quasi-concave, increasing in x, and nonincreasing in ℓ. Second-period utility is not discounted.

At the beginning of each period, a policy maker selects the parameters of a linear income tax:[7] the *tax rate* $t \in [0, 1]$ and the *guarantee* $T \in \Re$. With period τ tax system (t_τ, T_τ), a citizen with income y has a tax bill of $t_\tau y - T_\tau$. The first-period policy maker must also choose whether or not to undertake a discrete public investment. This decision is denoted by g, with $g = 1$ (0) meaning that the investment is (not) undertaken. The investment costs C units of the consumption good but is productive, raising all individuals' second-period abilities. Thus, for all citizens i, $a_{2i} = a_{2i}(g)$ where $a_{2i}(1) \geq a_{2i}(0)$.[8]

Labor supply decisions are made after the policy maker has selected the tax system. With tax system (t_τ, T_τ), citizen i will supply $\ell(t_\tau, T_\tau, a_{\tau i})$ units of labor, where $\ell(t, T, a) = \arg\max\{u((1-t)a\ell + T, \ell) : \ell \in [0, 1]\}$. He will earn $y(t_\tau, T_\tau, a_{\tau i})$ units of income, where $y(t, T, a) = a\ell(t, T, a)$, and will enjoy a utility level $v(t_\tau, T_\tau, a_{\tau i})$, where $v(t, T, a) = u((1-t)y(t, T, a) + T, \ell(t, T, a))$.

The policy maker's tax and investment choices must be feasible. In period one this amounts to satisfying the constraint:

$$(1) \quad t_1 \sum_{i=1}^{N} y(t_1, T_1, a_{1i}) - NT_1 = Cg.$$

The left-hand side of (1) is net tax revenues and the right-hand side is public expenditures. Let Z_1 denote the set of period-one policy choices (t_1, T_1, g) which satisfy (1). Since there are no public expenditures in period two, the policy maker's feasibility constraint is

$$(2) \quad t_2 \sum_{i=1}^{N} y(t_2, T_2, a_{2i}(g)) - NT_2 = 0.$$

Let $Z_2(g)$ denote the set of period-two policy choices (t_2, T_2) which satisfy (2), when the investment decision is g.

B. *The Political Process*

The model of policy-making is based on Besley and Coate (1997).[9] In each period a

geneous policy preferences. They also do not explain how the incumbent party came to be in power. Tabellini and Alesina (1990 p. 39) assume that decisions are made by majority rule. However, they need to impose structure on preferences to guarantee the existence of a median voter and must assume that the identity of the median voter changes over time, "due to: (i) random shocks to the costs of voting that affect the participation rate ...; or (ii) changes in the eligibility of the voting population."

[7] This could be motivated, following Geoffrey Brennan and Buchanan (1980), as a constitutional constraint on redistributive instruments. Our basic analytical method could be applied with any given set of redistributive instruments.

[8] The public investment may create incentives for individuals to borrow from one another: a citizen who expects to benefit greatly from the investment may gain from borrowing from a nonbeneficiary in the first period. To focus directly on the efficiency of public choices, however, we avoid the complications associated with introducing a private loan market.

[9] The Downsian model of policy-making with two-vote maximizing parties is difficult to apply to the environment

member of the community is selected to make policy choices, with an election determining the choice of citizen to do this. All citizens are able to run in these elections and each must choose whether or not to declare himself as a candidate. Running for office is not costly.[10] All individuals in the society then vote over the set of self-declared candidates. The candidate with the most votes wins (there is plurality rule). In the event of ties, the winning candidate is chosen randomly with each tying candidate having an equal chance of being selected. If only one individual runs for office then he is automatically selected. If no citizen runs, taxes are zero and laissez-faire prevails. The winning candidate selects policy only in the period for which he is elected.

The political process in each period is modeled as a three-stage game. Stage one sees each citizen deciding whether or not to become a candidate. In stage two, citizens vote over the set of self-declared candidates. At stage three, the winning candidate chooses the policy for that period. In voting, citizens anticipate candidates' policy choices and vote accordingly. As potential candidates, citizens anticipate this voting behavior. The two periods are linked via the public investment decision, which affects citizens' second-period abilities and hence their preferences over period-two policies. Citizens, as voters and policy makers, must anticipate the consequences of their first-period decisions for second-period outcomes.[11]

Viewed *ex ante,* the political process generates a probability distribution over policy sequences. A *policy sequence* is a pair $\{(t_1, T_1, g), \pi_2\}$ consisting of a first-period policy choice $(t_1, T_1, g) \in Z_1$ and a probability distribution over second-period policy choices $\pi_2 : Z_2(g) \to [0, 1]$. With policy sequence $\{(t_1, T_1, g), \pi_2\}$, the first-period choice is (t_1, T_1, g) and the second-period policy choice is $(t_2, T_2) \in Z_2(g)$ with probability $\pi_2(t_2, T_2)$. The probability distribution π_2 is generated by the political equilibrium arising in period two when the public investment decision is g.[12] Let $V_i((t_1, T_1, g), \pi_2)$ denote citizen i's expected utility under the policy sequence $\{(t_1, T_1, g), \pi_2\}$, defined as

$$(3)\quad V_i((t_1, T_1, g), \pi_2) = v(t_1, T_1, a_{1i}) + \sum_{(t_2,T_2) \in \Delta(\pi_2)} \pi_2(t_2, T_2) v(t_2, T_2, a_{2i}(g)),$$

where we use $\Delta(\cdot)$ to denote the *support* of the probability distribution; i.e., the set of outcomes which are selected with positive probability.

II. Equilibrium Policy Sequences

This section provides a detailed description of political decision-making in both periods, defining political equilibrium and characterizing equilibrium policy sequences. We begin with the second-period election and policy choice, taking as given the public investment decision. We then analyze the first-period election and policy choice, recognizing that citizens as voters and elected policy makers will anticipate the dependence of second-period policy choices on the public investment decision.

A. *Period-Two Election and Policy Choice*

We work through the three stages of the political process in reverse order, beginning with the policy-selection stage.[13] The citizen who wins the period-two election will implement his preferred period-two policy—promises to

of this paper since the first-period policy choice involves both a tax rate and an investment decision. Without further restriction, this standard model typically fails to produce a prediction in two-dimensional models.

[10] The results of the paper would be unaffected by the assumption that running for office has a small cost (say, in the form of reduced consumption).

[11] The model assumes that citizens are farsighted. While assuming myopic voting is common in dynamic public choice models, it is not very satisfactory when questions of efficiency are on the agenda. For a useful review of the different "shortcuts" taken in dynamic public choice models, see Per Krusell et al. (1997).

[12] Using a probability distribution over period-two policy choices captures any randomness in political outcomes stemming, for example, from the possibility of ties. The same randomness is possible in period one, which is why the political process generates a probability distribution over policy sequences.

[13] The treatment will be concise and the reader is referred to Besley and Coate (1997) for further details.

do otherwise are not credible. Citizen i's preferred policy is given by:

$$(4)\quad (t_{2i}(g), T_{2i}(g)) = \arg\max\{v(t_2, T_2, a_{2i}(g)) \mid (t_2, T_2) \in Z_2(g)\}.$$

We assume that the solution to (4) is unique.[14] Associated with each citizen's election, therefore, is a utility imputation $(v^i_{21}(g), \ldots, v^i_{2N}(g))$, where $v^i_{2j}(g) = v(t_{2i}(g), T_{2i}(g), a_{2j}(g))$ is individual j's period-two utility if i is elected. If no citizen stands for office, the second-period policy is $(t_{20}(g), T_{20}(g)) = (0, 0)$. We denote the utility imputation in this case as $(v^0_{21}(g), \ldots, v^0_{2N}(g))$.

We now turn to the voting stage. Let the set of candidates be $C \subset \mathcal{N}$ and let $\alpha_{2j} \in C \cup \{0\}$ denote citizen j's decision, where $\alpha_{2j} = i$ denotes j voting for candidate i, and $\alpha_{2j} = 0$ denotes abstention. A vector of voting decisions is $\boldsymbol{\alpha}_2 = (\alpha_{21}, \ldots, \alpha_{2N})$. Given C and $\boldsymbol{\alpha}_2$, let $P^i(C, \boldsymbol{\alpha}_2)$ denote the probability that candidate i wins. Under the assumption of plurality rule, $P^i(C, \boldsymbol{\alpha}_2)$ equals 1 if i has the most votes or is the only candidate. It equals $1/M$, if there are M tying candidates of which i is one. It is zero otherwise.

Individuals vote to maximize their expected utility given the voting decisions of others. A period-two *voting equilibrium* given g is a vector of decisions $\boldsymbol{\alpha}^*_2$ such that, for each citizen $j \in \mathcal{N}$: (i) α^*_{2j} is a best response to $\boldsymbol{\alpha}^*_{2-j}$, i.e.,

$$(5)\quad \alpha^*_{2j} \in \arg\max\left\{\sum_{i \in C} P^i(C, (\alpha_{2j}, \boldsymbol{\alpha}^*_{2-j})) \times v^i_{2j}(g) \mid \alpha_{2j} \in C \cup \{0\}\right\},$$

and (ii) α^*_{2j} is not a weakly dominated voting strategy. Ruling out weakly dominated strategies is standard in the voting literature and implies that citizens do not vote for their least preferred candidate. Hence voting is sincere in two-candidate elections.

We are now ready to describe the entry stage. Citizens run for office to move policy in their preferred direction, with their ability to do so dependent on who else enters. Let $s_{2j} \in \{0, 1\}$ denote citizen j's pure strategy, where $s_{2j} = 1$ denotes entry. Let $\mathbf{s}_2 = (s_{21}, \ldots, s_{2N})$ denote a pure strategy profile. The set of period-two candidates with strategy profile $\mathbf{s}_2$ is $C(\mathbf{s}_2) = \{i \mid s_{2i} = 1\}$.

Each citizen's expected payoff from any given strategy profile depends upon anticipated voting behavior. We let $\boldsymbol{\alpha}_2(C)$ be the voting behavior citizens anticipate when the candidate set is C. Thus, if citizen i enters then he/she anticipates winning with probability $P^i(C(\mathbf{s}_2), \boldsymbol{\alpha}_2(C(\mathbf{s}_2)))$. The expected payoff to any citizen j given $\mathbf{s}_2$ is therefore

$$(6)\quad U_{2j}(\mathbf{s}_2; g, \boldsymbol{\alpha}_2(\cdot)) = \sum_{i \in C(\mathbf{s}_2)} P^i(C(\mathbf{s}_2), \boldsymbol{\alpha}_2(C(\mathbf{s}_2)))v^i_{2j}(g) + P^0(C(\mathbf{s}_2))v^0_{2j}(g),$$

where $P^0(C(\mathbf{s}_2))$ denotes the probability that the default outcome is selected. This equals one if $C(\mathbf{s}_2) = \varnothing$ and zero otherwise. Citizen j's payoff is thus the probability that each candidate i wins in the race times his payoff from i's preferred policy.

To ensure the existence of an equilibrium, we need to allow for mixed strategies. Let γ_{2j} $(\in [0, 1])$ be a mixed strategy for citizen j, with a mixed-strategy profile being denoted by $\boldsymbol{\gamma}_2 = (\gamma_{21}, \ldots, \gamma_{2N})$ and citizen j's expected payoff under $\boldsymbol{\gamma}_2$ being denoted by $\omega_{2j}(\boldsymbol{\gamma}_2; g, \boldsymbol{\alpha}_2(\cdot))$. An *equilibrium of the period-two election* given g is a vector of entry decisions $\boldsymbol{\gamma}_2$ and a voting function $\boldsymbol{\alpha}_2(\cdot)$ such that: (i) for each citizen j, γ_{2j} is a best response to $\boldsymbol{\gamma}_{2-j}$

[14] While it is highly unlikely that (4) will have multiple solutions, it is difficult to come up with a set of simple sufficient conditions to rule it out. The conditions used by Meltzer and Richard (1981) to guarantee the inverse relationship between tax rates and productive ability ($u(\cdot)$ is strictly concave and satisfies the Inada conditions and both consumption and leisure are normal goods) are certainly sufficient, but much stronger than necessary.

given g and $\boldsymbol{\alpha}_2(\cdot)$, and (ii) for all nonempty candidate sets $C \subset \mathcal{N}$, $\boldsymbol{\alpha}_2(C)$ is a voting equilibrium given g.

Associated with any such equilibrium $(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot))$ are two probability distributions. First, a distribution over the citizen who chooses period-two policy denoted by $r_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot))$, where $r_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot))(i)$ is the probability that citizen i will be the policy maker and $r_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot))(0)$ the probability that nobody runs. Second, a distribution over second-period policies, denoted by $\pi_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot), g)$, where $\pi_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot), g)(t_2, T_2)$ is the probability that the second-period policy is (t_2, T_2). These probability distributions are related by

$$(7) \qquad \pi_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot), g)(t_2, T_2) = \sum_{j \in \{i \in \mathcal{N} \cup \{0\} : (t_{2i}(g), T_{2i}(g)) = (t_2, T_2)\}} r_2(\boldsymbol{\gamma}_2, \boldsymbol{\alpha}_2(\cdot))(j).$$

B. *Period-One Election and Policy Choice*

Citizens look forward in making their first-period decisions. If they anticipate an equilibrium of the period-two election $(\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot))$ when the investment decision is $g \in \{0, 1\}$, then the probability distribution over second-period policy choices is expected to be $\pi_2^g \equiv \pi_2(\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot), g)$, with $\Pi_2 \equiv (\pi_2^0, \pi_2^1)$. If elected as first-period policy maker, citizen i will select period-one policy

$$(8) \qquad (t_{1i}(\Pi_2), T_{1i}(\Pi_2), g_i(\Pi_2)) = \arg\max\{V_i((t_1, T_1, g), \pi_2^g) \mid (t_1, T_1, g) \in Z_1\},$$

given Π_2. We again assume that the solution to (8) is unique. Associated with each citizen's election is a utility imputation $(v_{11}^i(\Pi_2), \ldots, v_{1N}^i(\Pi_2))$, where $v_{1j}^i(\Pi_2) = V_j(t_{1i}(\Pi_2), T_{1i}(\Pi_2), g_i(\Pi_2); \pi_2^{g_i})$ is individual j's period-one expected utility if i is elected. If no citizen stands for office, the first-period policy is $(t_{10}(\Pi_2), T_{10}(\Pi_2), g_0(\Pi_2)) = (0, 0, 0)$ and the utility imputation is $(v_{11}^0(\Pi_2), \ldots, v_{1N}^0(\Pi_2))$.

The period-one voting stage parallels that in period two. Let the set of candidates be $C \subset \mathcal{N}$ and let $\alpha_{1j} \in C \cup \{0\}$ denote j's voting decision. A vector of voting decisions $\boldsymbol{\alpha}_1^* = (\alpha_{11}^*, \ldots \alpha_{1N}^*)$ is a period-one *voting equilibrium*, given Π_2, if for each citizen $j \in \mathcal{N}$: (i) α_{1j}^* is a best response to $\boldsymbol{\alpha}_{1-j}^*$, and (ii) α_{1j}^* is not a weakly dominated voting strategy.

Turning to candidate entry, $s_{1j} \in \{0, 1\}$ is each citizen's pure strategy, with $\alpha_1(\cdot)$ being the (common) function describing anticipated voting behavior. The payoff to any citizen j given his own entry decision and those of other citizens is given by

$$(9) \qquad U_{1j}(\mathbf{s}_1; \Pi_2, \boldsymbol{\alpha}_1(\cdot)) = \sum_{i \in C(\mathbf{s}_1)} P^i(C(\mathbf{s}_1), \boldsymbol{\alpha}_1(C(\mathbf{s}_1)))v_{1j}^i(\Pi_2) + P^0(C(\mathbf{s}_1))v_{1j}^0(\Pi_2).$$

Let γ_{1j} be a mixed strategy for citizen j and let $\boldsymbol{\gamma}_1 = (\gamma_{11}, \ldots, \gamma_{1N})$ be a mixed-strategy profile. Citizen j's expected payoff under the mixed strategy profile $\boldsymbol{\gamma}_1$ is denoted $\omega_{1j}(\boldsymbol{\gamma}_1, \Pi_2, \boldsymbol{\alpha}_1(\cdot))$.

An *equilibrium of the period-one election*, given Π_2, is a vector of entry decisions $\boldsymbol{\gamma}_1$ and a voting function $\boldsymbol{\alpha}_1(\cdot)$ such that: (i) for each citizen j, γ_{1j} is a best response to $\boldsymbol{\gamma}_{1-j}$ given Π_2, and (ii) for all nonempty candidate sets $C \subset \mathcal{N}$, $\boldsymbol{\alpha}_1(C)$ is a voting equilibrium given Π_2. Associated with any equilibrium $(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}_1(\cdot))$ are probability distributions over first-period policy makers $(r_1(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}_1(\cdot)))$ and over first-period policies $(\pi_1(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}_1(\cdot), \Pi_2))$.

C. *Political Equilibrium and Equilibrium Policy Sequences*

A *political equilibrium* is a triple $\{(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}_1(\cdot)), (\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot))_{g \in \{0,1\}}\}$ such that: (i) $(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}_1(\cdot))$ is an equilibrium of the period-one election given the probability distribution over second-period policies implied by $(\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot))_{g \in \{0,1\}}$, and (ii) $(\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot))$ is an equilibrium of the period-two election given g for $g \in \{0, 1\}$. An equilibrium policy sequence is one which could potentially arise in political equilibrium. More precisely, $\{(t_1^*, T_1^*, g^*), \pi_2^*\}$ is an *equilibrium policy sequence* if there exists a political equilibrium $\{(\boldsymbol{\gamma}_1, \boldsymbol{\alpha}_1(\cdot)), (\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot))_{g \in \{0,1\}}\}$ such that (t_1^*, T_1^*, g^*) lies in the support of the probability distribution over

first-period policies generated by that political equilibrium and π_2^* is the probability distribution over second-period policy choices generated by the political equilibrium when the public investment decision is g^*. It is straightforward to show that a political equilibrium, and hence an equilibrium policy sequence, exists.[15]

III. Analysis of Equilibrium Policy Sequences

This section investigates the properties of the policy sequences emerging from political equilibrium. We begin by noting that a certain type of efficiency property follows immediately from the fact that policy makers are utility-maximizing citizens.

PROPOSITION 1: *Let $\{(t_1^*, T_1^*, g^*), \pi_2^*\}$ be an equilibrium policy sequence and let $\{(\gamma_1, \alpha_1(\cdot)), (\gamma_2^g, \alpha_2^g(\cdot))_{g \in \{0,1\}}\}$ be the associated political equilibrium. Then, (i) if $\gamma_{1j} = 1$ for some $j \in \mathcal{N}$, there is no $(t_1, T_1, g) \in Z_1$ such that*

$$V_i((t_1, T_1, g), \pi_2(\gamma_2^g, \alpha_2^g(\cdot), g))$$

$$> V_i((t_1^*, T_1^*, g^*), \pi_2(\gamma_2^{g^*}, \alpha_2^{g^*}(\cdot), g^*))$$

$$\text{for all } i \in \mathcal{N},$$

and (ii) if $(t_2^*, T_2^*) \in \Delta(\pi_2^*)$, there is no $(t_2, T_2) \in Z_2(g^*)$ such that

$$v(t_2, T_2, a_{2i}(g^*)) > v(t_2^*, T_2^*, a_{2i}(g^*))$$

$$\text{for all } i \in \mathcal{N}.$$

Part (i) says that if at least one citizen enters the first-period election with probability one then, conditional on second-period policy choices being chosen through the democratic process and the equilibria of the second-period election being $(\gamma_2^g, \alpha_2^g(\cdot))_{g \in \{0,1\}}$, there exists no alternative first-period policy choice which gives every citizen more expected utility. This follows from the fact that the first-period policy choice is optimal for some first-period policy maker. Part (ii) says that there is no alternative second-period policy choice which, given the public investment decision, gives every citizen more second-period utility. The caveat that at least one citizen run is not necessary, since the default outcome is Pareto efficient in the second period.

Proposition 1 parallels a result (Proposition 10) in Besley and Coate (1997), and is a natural consequence of including the policy maker's utility in assessing efficiency. However, it does not imply that equilibrium policy sequences are efficient in the usual economists' sense. Standard notions of economic efficiency are defined without reference to the policy-making process, taking no account of the institutional constraints that shape actual decisions. In this vein, a policy sequence is efficient if there exists no other technologically feasible policy sequence which makes all citizens better off. There is no guarantee that policy sequences are efficient in this sense.

We illustrate this by focusing on the public investment decision. We suppose that the public investment is *potentially Pareto improving* in the sense that for any policy sequence $\{(t_1, T_1, 0), \pi_2\}$ there exists an alternative policy sequence $\{(t_1', T_1', 1), \pi_2'\}$ such that $V_i((t_1', T_1', 1), \pi_2') > V_i((t_1, T_1, 0), \pi_2)$ for all $i \in \mathcal{N}$. We will then identify three reasons why the investment might not be undertaken in political equilibrium.

We begin with a benchmark result which establishes a sufficient condition for the investment to be implemented in political equilibrium.

PROPOSITION 2: *Let $\{(t_1^*, T_1^*, g^*), \pi_2^*\}$ be an equilibrium policy sequence and let $\{(\gamma_1, \alpha_1(\cdot)), (\gamma_2^g, \alpha_2^g(\cdot))_{g \in \{0,1\}}\}$ be the associated political equilibrium. Suppose that*

[15] Proposition 1 of Besley and Coate (1997) implies that there exists at least one equilibrium of the period-two election for each possible level of g. Any pair of such equilibria (one for both levels of g) generates a pair of probability distributions over second-period policy choices. Proposition 1 of Besley and Coate (1997) also implies that there exists an equilibrium of the period-one election for any pair of probability distributions over second-period policy choices. Hence, there must exist a political equilibrium. Existence of equilibrium in this type of "citizen-candidate" model is not problematic because potential candidates face a discrete "in-or-out" decision, rather than a decision about which platform to adopt.

for all $j \in \Delta(r_1(\gamma_1; \alpha_1(\cdot)))$ *and* $k \in \Delta(r_2(\gamma_2^1, \alpha_2^1(\cdot)))$,

$$(t_{2j}(1), T_{2j}(1)) = (t_{2k}(1), T_{2k}(1)).$$

Then, if $\gamma_{1j} = 1$ *for some* $j \in \mathcal{N}$ *and the public investment is potentially Pareto improving,* $g^* = 1$.

PROOF:
See Appendix.

The key condition is that all first-period policy makers anticipate that, were they to undertake the investment, the second-period policy maker would choose exactly the same policy as they would. If the investment is potentially Pareto improving, there exists (by definition) a way of redistributing the gains so that all individuals benefit. Thus a policy maker who knows that someone who shares his policy preferences will govern next period is sure that the redistribution, as needed to make the investment worthwhile for him, will be forthcoming. The following example shows how political instability can lead the current incumbent to forgo the investment.

Example 1: Suppose that the population is equally divided into two ability types—low and high. Let $\mathcal{N}_L$ denote the set of low-ability individuals and $\mathcal{N}_H$ the set of high-ability individuals. Assume that individuals care only about their consumption and are risk neutral, so that $u(x, \ell) = x$. Let a_L denote the first-period ability of low-ability individuals and a_H $(>a_L)$ that of high individuals. If the investment is not undertaken, individuals' second-period abilities equal those in the first period. If it is undertaken, then it raises the ability of high-ability individuals. Formally, $a_{2i}(1) = a_H + \delta$ for $i \in \mathcal{N}_H$, while $a_{2i}(1) = a_L$ for $i \in \mathcal{N}_L$. We assume that $\delta/2 > C/N$, which implies that the public investment is potentially Pareto improving.

We now demonstrate the existence of an equilibrium policy sequence in which the investment is not undertaken. We begin by describing the equilibria of the second-period election. Note first that if a high-ability citizen is elected to office in period two, he will choose zero taxation irrespective of the public investment decision, implying that $(t_{2H}(1), T_{2H}(1)) = (t_{2H}(0), T_{2H}(0)) = (0, 0)$. A low-ability individual will choose a tax rate of 100 percent and hence $(t_{2L}(1), T_{2L}(1)) = (1, \bar{a} + \delta/2)$ and $(t_{2L}(0), T_{2L}(0)) = (1, \bar{a})$, where $\bar{a} = (a_L + a_H)/2$ is the mean first-period ability.

Irrespective of the public investment decision, therefore, there is an equilibrium of the second-period election in which a low-ability individual runs against a high-ability individual and all citizens vote for candidates who share their ability. (It is simple to check that entry is worthwhile and that no other candidate will enter.) In this equilibrium, each candidate is elected with probability 1/2. When $g = 0$, the probability distribution over second-period policies generated by this equilibrium selects $(1, \bar{a})$ with probability 1/2 and $(0, 0)$ with probability 1/2. When $g = 1$, the probability distribution selects $(1, \bar{a} + \delta/2)$ with probability 1/2 and $(0, 0)$ with probability 1/2.

Turning to the period-one election, observe that a low-ability citizen will set a tax rate of 100 percent if he is elected; the only question is whether he will undertake the investment. It is simple to show that he will not introduce the investment if $\delta/4 < C/N$. Under this assumption, therefore, a low-ability citizen will select the first-period policy $(1, \bar{a}, 0)$. A high-ability citizen would always set a zero tax rate and would always undertake the investment, thereby selecting $(0, -C/N, 1)$ as the first-period policy. It follows that there is an equilibrium of the period-one election in which a single low-ability individual runs against a single high-ability individual and all citizens vote for candidates who share their ability. In this equilibrium, each candidate is elected with probability 1/2. Assuming that $\delta/4 < C/N$, the probability distribution over first-period policies generated by this equilibrium selects $(1, \bar{a}, 0)$ with probability 1/2 and $(0, -C/N, 1)$ with probability 1/2.

Hence, if $\delta/4 < C/N$, there is an equilibrium policy sequence involving the period-one policy choice $(1, \bar{a}, 0)$ and a probability distribution over second-period policies selecting $(1, \bar{a})$ with probability 1/2 and $(0, 0)$ with probability 1/2.

In this example, the investment harms the first-period policy maker in period one,

quiring period-two compensation if he is to be induced to undertake it. Such compensation will be forthcoming if an individual with similar preferences is elected to office, but not otherwise. The possibility that an individual with opposing preferences will be elected thus deters him from undertaking the investment.[16] In this way, political turnover can induce policy makers to resist making public investments which entail short-run costs, if the future gains are contingent on actions of future policy makers. This is a general problem which is likely to crop up in many public decisions. Any policy with short-run costs and long-run benefits which do not accrue uniformly may not be undertaken for this reason.

We now turn to public investments where, unlike in this example, compensations by manipulating taxes and transfers are not necessary to achieve Pareto gains—all citizens benefit from the investment, provided only that the tax rate does not increase and the guarantee does not decrease. In pursuit of a formal definition, we first develop the idea of one probability distribution over second-period policies dominating another because it involves lower tax rates and higher guarantees. Thus, for any $(t_2, T_2) \in Z_2(0)$, let $\Psi(t_2, T_2)$ denote the set of second-period policy pairs in $Z_2(1)$, for which the income guarantee is at least as high and the tax rate no higher. We say that the probability distribution $\pi_2' : Z_2(1) \to \Re_+$ *dominates* the probability distribution $\pi_2 : Z_2(0) \to \Re_+$ if: (i) for every $(t_2', T_2') \in \Delta(\pi_2')$ there exists $(t_2, T_2) \in \Delta(\pi_2)$ such that $(t_2', T_2') \in \Psi(t_2, T_2)$, and (ii) for all $(t_2, T_2) \in \Delta(\pi_2)$,

$$\pi_2(t_2, T_2) \leq \sum_{(t_2', T_2') \in \Psi(t_2, T_2)} \pi_2'(t_2', T_2').$$

The first part of the definition says that every second-period policy pair selected by π_2' involves a lower tax rate and a higher guarantee than *some* policy pair selected by π_2. The second says that, for every second-period policy pair selected by π_2, the dominant distribution π_2' selects a policy pair involving a lower tax rate and a higher guarantee with equal or greater probability. If the investment is undertaken in period one and the associated probability distribution over second-period policies dominates that which would have occurred absent the investment, the expected utility of any citizen i in period two cannot be lowered. As a final preliminary, let $Z_2^*(g)$ $(\subset Z_2(g))$ denote the set of period-two policy choices which are not strictly Pareto dominated.[17]

Now we define the public investment to be *straightforwardly Pareto improving* if for every policy sequence $\{(t_1, T_1, 0), \pi_2\}$ with $\Delta(\pi_2) \subset Z_2^*(0)$ there exists $(t_1', T_1', 1) \in Z_1$ such that if π_2' dominates π_2, then $V_i((t_1', T_1', 1), \pi_2') > V_i((t_1, T_1, 0), \pi_2)$ for all $i \in \mathcal{N}$. This definition requires that for any given policy sequence in which the investment is not undertaken, there exists a way of financing the investment which makes all citizens better off for all dominating probability distributions over second-period policy pairs.[18] The requirement that the investment be Pareto improving for every dominating distribution is the difference with the earlier definition and embodies the idea of no redistribution being necessary to achieve an improvement.

In Example 1 the investment was not straightforwardly Pareto improving. Consider the policy sequence involving a first-period policy choice $(1, \bar{a}, 0)$ and a second-period policy choice $(0, 0)$ with probability one. Any dominating probability distribution over $Z_2(1)$ must also select $(0, 0)$ with probability one.

[16] A related argument is made by Avinash Dixit and John Londregan (1995), who argue that efficient private-sector decisions can be deterred because of the inability of political parties to commit to let individuals suffer economic losses and/or keep economic gains.

[17] Thus, $(t_2, T_2) \in Z_2^*(g)$ if and only if there does not exist $(t_2', T_2') \in Z_2(g)$ such that $v(t_2', T_2', a_{2i}(g)) > v(t_2, T_2, a_{2i}(g))$ for all $i \in \mathcal{N}$.

[18] The definition also restricts comparisons to feasible policy sequences which select only Pareto-undominated tax rates in period two [i.e., which satisfy $\Delta(\pi_2) \subset Z_2^*(0)$]. This, for example, rules out comparisons with policy sequences which set t_2 so high that individuals are deterred from supplying any labor in period two. In that case, no investment, no matter how much it enhanced productivity, could satisfy the requirements of the definition.

Hence, there is no way of financing the investment in the first period to make low-ability individuals better off for all dominating probability distributions.

It would seem more hopeful that straightforwardly Pareto-improving investments would be chosen in political equilibrium. The next result gives some conditions for this to be the case.

PROPOSITION 3: *Let* $\{(t_1^*, T_1^*, g^*), \pi_2^*\}$ *be an equilibrium policy sequence and let* $\{(\gamma_1, \boldsymbol{\alpha}_1(\cdot)), (\gamma_2^g, \boldsymbol{\alpha}_2^g(\cdot))_{g \in \{0,1\}}\}$ *be the associated political equilibrium. Suppose that: (i)*

$$r_2(\gamma_2^0, \boldsymbol{\alpha}_2^0(\cdot)) = r_2(\gamma_2^1, \boldsymbol{\alpha}_2^1(\cdot)),$$

and (ii) for all $i \in \Delta(r_2(\gamma_2^1, \boldsymbol{\alpha}_2^1(\cdot)))$,

$$(t_{2i}(1), T_{2i}(1)) \in \Psi(t_{2i}(0), T_{2i}(0)).$$

Then, if $\gamma_{1j} = 1$ *for some* $j \in \mathcal{N}$ *and the public investment is straightforwardly Pareto improving,* $g^* = 1$.

PROOF:
See Appendix.

This proposition weakens the condition of complete political stability from Proposition 2, requiring only that the investment does not affect the probability distribution over future policy makers or cause them to select a policy pair involving a higher tax rate or lower guarantee. Both of these conditions were satisfied in Example 1. Nonetheless, these conditions are strong. Our next two examples illustrate cases where conditions (i) and (ii) of the proposition fail and straightforwardly Pareto-improving investments are not undertaken. The examples are very stylized, intending to give pure cases of each violation.

Example 2: The population is divided into three groups of individuals: high types, low types and movers. Let $\mathcal{N}_H$ denote the set of high types; $\mathcal{N}_L$ the set of low types and $\mathcal{N}_M$ the set of movers. Assume that $\#\mathcal{N}_L + \#\mathcal{N}_M = \#\mathcal{N}_H$ and that $\#\mathcal{N}_M < N/3$. Preferences are $u(x, \ell) = x$. High and low types have abilities a_H and a_L in both periods, where $a_H > a_L$. Thus, neither of these groups receives a direct benefit from the public investment. Movers have ability a_L in the first period but, if the investment is undertaken, have ability $a_L + \delta$ in period two. We assume that the gain to the movers is such that $N(a_H - a_L)/2(N - \#\mathcal{N}_M) < \delta < 3(a_H - a_L)/4$. In addition, suppose that the investment is costless ($C = 0$), which implies that the investment is straightforwardly Pareto improving.[19]

Again we will demonstrate the existence of an equilibrium policy sequence in which the investment is not undertaken. Beginning with period two, a low-type policy maker would set a tax rate of 100 percent whether or not the investment is undertaken, i.e., $(t_{2L}(0), T_{2L}(0)) = (1, \bar{a})$ and $(t_{2L}(1), T_{2L}(1)) = (1, \bar{a} + (\#\mathcal{N}_M/N)\cdot\delta)$ where $\bar{a} = (a_L + a_H)/2$. A high-ability policy maker would choose zero taxation, so that $(t_{2H}(g), T_{2H}(g)) = (0, 0)$ for $g \in \{0, 1\}$. If the investment is not undertaken, the movers share low types' preferences and hence $(t_{2M}(0), T_{2M}(0)) = (1, \bar{a})$. However, if the first-period investment is made, then under our assumptions on δ, movers do not want any redistributive taxation implying that $(t_{2M}(1), T_{2M}(1)) = (0, 0)$.

It follows that if the investment is not undertaken, there is an equilibrium of the second-period election in which a low-ability candidate and a high-ability candidate tie. The low-ability candidate is supported by both low-ability citizens and movers, while high-ability citizens vote for the high-ability candidate. The probability distribution over second-period policies is $(1, \bar{a})$ with probability $1/2$ and $(0, 0)$ with probability $1/2$. When the investment is undertaken, however, there is an equilibrium of the second-period election in

[19] This is not quite accurate, since when dealing with a policy sequence $\{(t_1, T_1, 0), \pi_2\}$ such that π_2 selects (0, 0) with probability one, it is only possible to find a $(t_1', T_1', 1)$ such that $\{(t_1', T_1', 1), \pi_2'\}$ weakly Pareto dominates $\{(t_1, T_1, 0), \pi_2\}$ if π_2' dominates π_2. This is because if π_2' dominates π_2, it must be the case that π_2' also selects (0; 0) with probability one. Thus the high- and low-ability types cannot both be made strictly better off, because they do not share any of the benefits of the investment. However, if we allowed the investment to also give a very small gain to the high- and low-ability citizens, then the desired condition would be satisfied even in this case and the analysis would be unchanged.

which a single high-ability candidate runs uncontested. No mover has an incentive to enter, since movers have the same policy preferences as high-ability citizens. Low-ability citizens would lose for sure if they entered. The probability distribution over second-period policies generated by this equilibrium selects (0, 0) with probability 1.

In period one, a low-ability policy maker would set a tax rate of 100 percent; the only question is whether he would undertake the investment. This depends on how it affects low types' second-period utility. If the investment is undertaken, a low type's expected second-period payoff is a_L, while if it is not undertaken, this payoff is $(3a_L + a_H)/4$. Thus a low-ability policy maker would choose $(1, \bar{a}, 0)$. A high-ability policy maker would select $(0, 0, 1)$, choosing no taxation and undertaking the investment. A mover would select $(1, \bar{a}, 1)$, setting a tax rate of 100 percent and also undertaking the investment.

It is now straightforward to find a two-candidate equilibrium of the first-period election involving a low type running against a high type. The low types and movers vote for the low-ability candidate, while the high-ability voters vote for the high-ability candidate. Anticipated voting behavior is such that if a mover entered the race, the high-ability candidate would be expected to win. To verify that movers prefer the low type to the high type, note that with a high type, a mover's lifetime utility is $2a_L + \delta$, while under a low-ability type it is $3a_H/4 + 5a_L/4$. The former is higher than the latter under our assumption on δ.[20]

Hence, there exists an equilibrium policy sequence with period-one policy choice $(1, \bar{a}, 0)$ and a probability distribution over second-period policies selecting $(1, \bar{a})$ with probability $1/2$ and $(0, 0)$ with probability $1/2$.

In this example, the first-period policy maker does not undertake the investment, even though nobody apparently loses from it, because it makes certain a loss of control. The beneficiaries of the investment switch allegiance and "gang up" on their former political allies, eliminating the possibility that a citizen who shares the first-period policy maker's preferences will be elected in period two. In this way, condition (i) of Proposition 3 is violated. The basic idea that current policy makers may distort current policies to influence the identity of future policy makers would appear quite general, with the papers by Aghion and Bolton (1990) and Milesi-Ferretti and Spolaore (1994) representing other possible applications.

Example 3: Suppose that the population is divided into three groups: poor, middle class, and rich. Let $\mathcal{N}_R$, $\mathcal{N}_M$, $\mathcal{N}_P$ denote the sets of rich, middle class, and poor individuals, respectively. Suppose that the number of poor citizens equals the number of rich and middle class citizens: i.e., $\#\mathcal{N}_P = \#\mathcal{N}_M + \#\mathcal{N}_R$. Assume that individuals' preferences take the form: $u(x, l) = x - l^2/2$. Let a_J denote the first-period ability of individuals in social class J ($J \in \{P, M, R\}$) and assume $a_P < a_M < a_R$. The public investment raises each middle class individual's ability by an amount $\delta \in (0, a_R - a_M)$, but does not affect poor or rich individuals' abilities. The public investment is again costless, which implies that it is straightforwardly Pareto improving.

We demonstrate the existence of a political equilibrium in which the investment is not undertaken. To solve for the second-period policies which would be selected by each type of citizen, observe that with this utility function, labor supply, and hence earnings, depend only on t. Hence, if $(t_2, T_2) \in Z_2(g)$, then $T_2 = t_2 \cdot \bar{y}(t_2, g)$ where $\bar{y}(t_2, g)$ is mean second-period earnings when the tax rate is t_2 and the investment decision is g. The optimization in (4) is thus simplified to $t_{2i}(g) = \arg\max v(t_2,$

[20] This example illustrates the difference between our model of the political process and the Downsian framework. Having two competing parties offer the first-period policies $(1, \bar{a}, 0)$ and $(0, 0, 1)$ would be inconsistent with equilibrium. The party offering the platform $(0, 0, 1)$ would change its policy to attract the votes of the movers. In our model, while the high-type candidate may have an incentive to promise more redistribution in the first period, such promises are not credible. The Downsian model circumvents this difficulty by assuming that parties just care about winning and thus can commit to implement any platform they propose.

$t_2 \cdot \bar{y}(t_2, g), a_{2i}(g))$. Using the envelope theorem, we obtain the following first-order condition:

$$\bar{y}(t_{2i}, g) - y(t_{2i}, a_{2i}(g)) + t_{2i} \cdot \partial\bar{y}(t_{2i}, g)/\partial t_2 \leq 0 \; (=0 \quad \text{if } t_{2i} > 0).$$

Since $y(t, a) = (1 - t)a^2$, the optimal tax rates of the three types of citizens can be solved for as

$$t_{2J}(g) = \max\left(0, \frac{A(g) - a_{2J}(g)^2}{(2A(g) - a_{2J}(g)^2)}\right)$$

$$J \in \{P, M, R\},$$

where $A(g) = [\#\mathcal{N}_P a_{2P}(g)^2 + \#\mathcal{N}_M a_{2M}(g)^2 + \#\mathcal{N}_R a_{2R}(g)^2]/N$. Thus a poor citizen chooses a positive tax rate, while a rich individual chooses zero taxation. A middle class individual might choose a zero or positive tax rate depending on the parameters.

Irrespective of the public investment decision, there is an equilibrium of the second-period election in which a poor individual runs and ties with a middle class individual. Poor citizens vote for the poor candidate, and the middle class candidate is supported by the rich and middle class. Anticipated voting behavior is such that in a three-way race between one candidate from each group, either the poor candidate would win or the poor and middle class candidate would tie, thus ensuring that no rich citizen has an incentive to enter. When the public investment decision is g, the probability distribution over second-period policies generated by this equilibrium selects $(t_{2M}(g), t_{2M}(g) \cdot \bar{y}(t_{2M}(g), g))$ with probability $1/2$ and $(t_{2P}(g), t_{2P}(g) \cdot \bar{y}(t_{2P}(g), g))$ with probability $1/2$.

In the period-one election, observe first that, since the investment is not costly, $(t_{1J}, T_{1J}) = (t_{2J}(0), T_{2J}(0))$ for all types $J \in \{P, M, R\}$. Whether the investment is undertaken depends on its effects on period-two policy choices. By raising the income-generating abilities of the middle class, the investment raises mean second-period earnings at any positive tax rate. This implies that any given tax rate produces a higher guarantee. It also raises the tax rate that would be chosen by a poor candidate and, provided that $t_{2M}(0) > 0$, reduces that chosen by a middle class candidate.

The middle class benefits directly from the investment in period two. While there is a cost to the middle class in terms of a higher tax rate if a poor candidate is elected, it can readily be shown that this is never sufficient to outweigh the direct benefits of the investment. Thus, a middle class policy maker would always invest. A rich policy maker must weigh up the increase in the tax rate chosen by a second-period poor candidate against the benefits of the higher mean second-period earnings and any reduction in the tax rate chosen by a middle class candidate. A poor policy maker weighs up the higher guarantee if the poor candidate is elected against a (possibly) lower guarantee if the middle class candidate is elected. The latter requires $t_{2M}(1) \cdot \bar{y}(t_{2M}(1), 1) < t_{2M}(0) \cdot \bar{y}(t_{2M}(0), 0))$. If this inequality is not satisfied, then the investment is unambiguously desirable from the viewpoint of a poor policy maker.

The fraction of middle class individuals in the economy is crucial to whether a poor candidate would wish to forgo the investment and a rich candidate would wish to undertake it. When $\#\mathcal{N}_M/N$ is small, the gains to the poor person in terms of raising mean earnings are small, but the costs in terms of less redistribution remain. For a rich policy maker, when $\#\mathcal{N}_M/N$ is small the costs in terms of an increase in the poor candidate's optimal second-period tax rate become small, but the benefits in terms of a reduction in the optimal tax rate of the middle class remain. This yields the following result.

LEMMA 1: *Suppose that* $(a_P^2 + a_R^2)/2 >$ a_M^2. *Then there exists* $\hat{\lambda} > 0$ *such that a poor first-period policy maker would forgo the investment and a rich first-period policy maker would undertake it if* $\#\mathcal{N}_M/N < \hat{\lambda}$.

PROOF:

See Appendix.

Thus, if $\#\mathcal{N}_M/N$ is small, there is an equilibrium of the first-period election involving a middle class candidate running against a poor

candidate. The rich and middle class vote for the middle class candidate, while the poor vote for the poor candidate. Each candidate is elected with probability $^1/_2$. The probability distribution over first-period policies generated by this equilibrium selects $(t_{2M}(0), t_{2M}(0)\cdot\bar{y}(t_{2M}(0), 0), 1)$ with probability $^1/_2$ and $(t_{2P}(0), t_{2P}(0)\cdot\bar{y}(t_{2P}(0), 0), 0)$ with probability $^1/_2$. In this case, there is an equilibrium policy sequence with period-one policy choice $(t_{2P}(0), t_{2P}(0)\cdot\bar{y}(t_{2P}(0), 0), 0)$ and second-period choices of $(t_{2P}(0), t_{2P}(0)\cdot\bar{y}(t_{2P}(0), 0))$ and $(t_{2M}(0), t_{2M}(0)\cdot\bar{y}(t_{2M}(0), 0))$ with equal probability.

In this example, the first-period policy maker does not undertake the investment because it changes the policy preferences of a future policy maker in a way that is detrimental to the incumbent. Specifically, it induces a future policy maker to prefer a lower guarantee, which violates condition (ii) of Proposition 2. Again, the basic idea that current policies might be manipulated to influence the choices of future leaders would appear quite general, with the paper by Persson and Svensson (1989) representing another application.

Together, the three examples reflect the importance of future policy for current choices. In the first example, future compensation is required to make the investment worthwhile for the incumbent. In the second and third, the endogeneity of future political outcomes drives the results. In these latter examples, it is key that the public investment alters citizens' preferences for future redistributive taxation. The arguments here are relevant more generally when current policies alter citizens' preferences over future policies. Thus, they are unlikely to be relevant for small projects. In order to apply them to public goods that generate future consumption benefits, we would require rather strong assumptions about preferences so that benefits could alter citizens' preferences over other types of policies.

All three examples assume that the public investment benefits a particular group disproportionately. This might suggest that programs with uniform benefits would be implemented efficiently. However, since political equilibrium tends not to be unique in a citizen-candidate model, no efficiency proposition is available for the uniform benefit case. This is because citizens' may anticipate a different future political equilibrium when an investment is undertaken. This may be true even if the distribution of benefits implies that a similar political equilibrium exists with the investment. Condition (i) of Proposition 3 may then fail because of the way in which citizens anticipate future political outcomes to change. While Example 2 appears to be a more compelling instance of the failure of this condition (not resting on multiple political equilibria), it is clear that the prognosis for general efficiency results in dynamic models of political equilibrium is even poorer than the above examples might suggest.

IV. Discussion

The results leave no doubt that public investments satisfying standard criteria of efficiency will not necessarily be adopted in political equilibrium. However, in view of the efficiency result in Proposition 1, it is not clear whether these should be viewed as "political failures" or whether technological definitions of efficiency should be questioned.

Despite the widespread use of the term "political failure" in public choice writings, it is hard to find any precise definition of this idea. It seems desirable to have a definition that parallels the usual definition of a market failure as closely as possible. Recall, therefore, the textbook analysis of market efficiency. First, the set of efficient allocations is characterized (graphically, the utility possibility frontier). This is a purely technological notion of efficiency, since the frontier depends only on the tastes and technologies of the economy. The second step requires a model, such as that developed by Arrow-Debreu, to specify how markets allocate resources. The idea of market failure then comes from observing that, under certain conditions, markets do not result in allocations that are on the frontier. The term "failure" is justified by the observation that, in principle, all citizens could be made better off.

To provide a parallel definition of political failure, one must similarly begin by defining the set of technologically feasible utility

allocations. This should reflect the available policy instruments. It is unreasonable, for example, to hold representative democracy to the standards of what might be achieved were lump-sum taxes and transfers available, while restricting policy makers to using only distortionary or "uniform" taxation.[21] At the second stage, political institutions are modeled, with a focus on whether equilibrium policy choices result in utility allocations on the frontier. By analogy with markets, a political failure arises when equilibrium policy choices leave the possibility of feasible, Pareto-improving policy choices.

By this standard, the nonimplementation of potentially Pareto-improving public investments described in the previous section represents a political failure; if such an investment is not implemented then there exists a Pareto-improving policy sequence. There is no conflict between this finding and Proposition 1, since that result says only that, conditional on second-period policies being selected through the democratic process, there exist no alternative first-period choices which can raise the expected utilities of all citizens. Hence, if a potentially Pareto-improving public investment is not undertaken, then it must be because the resulting *second-period equilibrium* choices would have made some citizens worse off. Thus, making future policy choices through the democratic process constrains society to the interior of the set of technologically feasible utility allocations.

We anticipate two kinds of objection to our claims of political failure. The first accepts our definition, but disputes the way in which we have analyzed the issue. The second takes issue with the definition. In the first category, we include a Coasian-style challenge to our analysis, which argues that bargaining between citizens and policy makers should be able to eliminate the inefficiency. It is straightforward to see that an incumbent policy maker would implement the investment in return for future policy makers setting second-period policies which would compensate them. By ignoring such gains from bargaining, our analysis might be regarded as misleading, just as Ronald Coase (1960) criticized the usual approach to externalities in a market context.

In response, we observe that the kind of bargain needed to restore efficient policy would require a significant amount of *commitment*: future policy makers would have to commit to the policies that they would select in the second period, and citizens would have to commit to elect these individuals in the future. Thus, the Coasian bargain would have to involve a significant fraction of the polity. It does not seem unreasonable to presume that transactions costs would hinder the undertaking of such bargains.[22]

A further objection to the analysis is that the political failures identified here could be eliminated if policy makers were elected for two periods. (This is immediate from Proposition 2.) However, our model provides no justification for holding a second-period election. While technically correct, this critique takes the model's structure too much at face value. In reality, democracy is characterized by periodic, rather than one-time, elections and our model captures this as simply as possible. Given that it is a commonly observed feature of democratic government, it is reasonable to suppose that periodic elections have a rationale. This suggests that any benefits from them exceed the social cost from the policy distortions identified here. However, this is different from arguing that the distortions that we have identified do not exist. They will arise in any system with periodic elections where government policies have effects that last beyond the time of the next election.

The main objection to our definition is the argument that it is insufficient to demonstrate a political failure by showing that there exist some technologically feasible policy choices

[21] It is common in the literature to overlook this point. Consider, for example, the textbook argument that representative democracy produces an inefficient level of public goods. The level of public goods demanded by the median voter is claimed to be "inefficient" because, in general, it does not satisfy the Samuelson rule. The latter is, however, derived under the assumption that lump-sum taxes and transfers are feasible.

[22] Indeed, one suspects that transactions costs are the main reason why societies resort to more delegated forms of decision-making (see Dixit [1996] for further discussion).

that produce a Pareto-superior utility allocation. Instead, it behooves the analyst to specify an alternative institutional arrangement for making policy decisions which actually selects such policy choices. Absent this, the term "failure" is misleading. Similar objections can, of course, be made about the notion of a market failure. In particular, there is a need to demonstrate that the government (or some other institution for allocating resources) could do better (see Buchanan [1962] and Oliver Williamson [1994] for discussions along these lines).

In response, it is important to make clear that we are not attempting to answer the broader question of whether representative democracy is desirable in the class of institutions that could be used to make collective choices (although answering this would constitute an interesting research agenda). Clearly, the ability to implement potentially Pareto-improving public investments represents only one dimension of this question. What is important is to reach a consistent position on markets and governments. Our definition is intended to achieve that. Just as the presence of market failures does not imply abandonment of the market system, the existence of political failures does not mean that representative democracy should be forsaken.[23] Indeed, the revealed preference for representative democracy, confirmed even more strongly in recent years, suggests that the benefits of democratic policy-making outweigh its costs.

There are, however, still important reasons to ask whether representative democracy results in efficient policy choices. First, it is central to many issues in public economics. In the absence of political failures, the prescriptions of normative economics (particularly those concerning efficiency) may be interpreted as predictions about actual policy choices. If observed policy choices do not correspond to these predictions, then, instead of berating policy makers for their foolishness, it is more constructive to provide explanations for their choices (a conclusion reminiscent of Stigler, 1982). The second reason is more practical: developing an understanding of why political failures occur may suggest efficiency-enhancing, piecemeal institutional reforms.

V. Concluding Remarks

The main achievements of this paper are twofold. First, it offers a precise definition of political failure. The notion put forward takes the standard technological definition of what is feasible for the citizens. This, we argued, is the most natural parallel to the literature on market failure. Second, the paper identifies three distinct classes of failures which arise when elections are repeated, even though citizens are forward looking and make optimizing decisions. Potentially Pareto-improving public investments may not be introduced because of fears that compensation needed to cover current costs will not be delivered and the fact that such projects may change the identities and/or policy choices of future leaders. These reasons can be traced to the fact that preferences over policy extend into the future, while political control does not.

We acknowledged that a Coasian argument could be levied against our claims of political failure; bargaining between citizens over current and future policies could eliminate the inefficiencies. However, this would require the ability to reach binding agreements about future policy, as well as a forum to bring the polity together. This vision is too utopian, given the transactions costs involved. More generally, however the issue of how institutions might evolve to minimize the economic costs associated with repeated elections is worthy of further study.

Much remains to be done to understand comprehensively the efficiency of resource allocation in a representative democracy, using the approach developed here and in our earlier paper. We have not yet modeled private-sector accumulation decisions and the effect of policy on these. Introducing imperfect

[23] In the market-failure literature, it is all too common to reach the conclusion that the existence of market failure suggests that the government could do better. This stems from the bad habit of viewing policy as being selected by a social planner. Showing that a planner could do better than the market is really no advance over pointing out that there exists a technologically feasible, Pareto-superior allocation. See Buchanan (1972) for further elaboration of this argument.

information also seems like an interesting avenue for future exploration. Finally, it would be interesting to incorporate other political institutions, such as pressure groups and legislative decision-making, into the analysis.

APPENDIX

PROOF OF PROPOSITION 2:

We know that for some citizen $j \in \mathcal{N}$, (t_1^*, T_1^*, g^*) solves the problem:

$$\text{Max } V_j((t_1, T_1, g), \pi_2(\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot), g))$$

$$\text{s.t. } (t_1, T_1, g) \in Z_1.$$

Suppose that $g^* = 0$. Then, since the investment is potentially Pareto improving, we know that there exists a policy sequence $\{(t_1', T_1', 1), \pi_2'\}$ such that $V_i((t_1', T_1', 1), \pi_2') > V_i((t_1^*, T_1^*, 0), \pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0))$ for all $i \in \mathcal{N}$. We will show that $V_j((t_1', T_1', 1), \pi_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot), 1)) > V_j((t_1^*, T_1^*, 0), \pi_2(\boldsymbol{\gamma}_2^0, \alpha_2^0(\cdot), 0))$, which will contradict the hypothesis that $g^* = 0$.

By hypothesis, for all $k \in \Delta(r_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot)))$,

$$(t_{2k}(1), T_{2k}(1)) = (t_{2j}(1), T_{2j}(1)).$$

Thus, $\pi_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot), 1)$ selects citizen j's optimal policy choice $(t_{2j}(1), T_{2j}(1))$ with probability one and hence

$$V_j((t_1', T_1', 1), \pi_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot), 1))$$

$$\geq V_j((t_1', T_1', 1), \pi_2')$$

$$> V_j((t_1^*, T_1^*, 0), \pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0)).$$

PROOF OF PROPOSITION 3:

We know that for some citizen $j \in \mathcal{N}$, (t_1^*, T_1^*, g^*) solves the problem:

$$\text{Max } V_j((t_1, T_1, g), \pi_2(\boldsymbol{\gamma}_2^g, \boldsymbol{\alpha}_2^g(\cdot), g))$$

$$\text{s.t. } (t_1, T_1, g) \in Z_1.$$

Suppose that $g^* = 0$. Then, since $\{(t_1^*, T_1^*, g^*), \pi_2^*\}$ is a political equilibrium policy sequence, Proposition 1 tells us that $\Delta(\pi_2^*) \subseteq Z_2^*(0)$. Thus, since the investment is straightforwardly Pareto improving, we know that there exists $(t_1', T_1', 1) \in Z_1$ such that if π_2' dominates π_2^*, $V_i((t_1', T_1', 1), \pi_2') > V_i((t_1, T_1, 0), \pi_2^*)$ for all $i \in \mathcal{N}$. We will show that

$$V_j((t_1', T_1', 1), \pi_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot), 1))$$

$$> V_j((t_1^*, T_1^*, 0), \pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0)),$$

which will contradict the hypothesis that $g^* = 0$. To do this it suffices to demonstrate that $\pi_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot), 1)$ dominates $\pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0)$.

Recalling the definition of dominance in Section III, there are two conditions to check. First, let $(t_2', T_2') \in \Delta(\pi_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot), 0))$. Then we know that there exists some $k \in \Delta(r_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot)))$ such that $(t_{2k}(1), T_{2k}(1)) = (t_2', T_2')$. By hypothesis, we have that $k \in \Delta(r_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot)))$ which implies that $(t_{2k}(0), T_{2k}(0)) \in \Delta(\pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0))$. In addition, again by hypothesis, $(t_{2k}(1), T_{2k}(1)) \in \Psi(t_{2k}(0), T_{2k}(0))$. This establishes the first condition.

To check the second condition, note that if $(t_2, T_2) \in \Delta(\pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0))$ then

$$\pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0)(t_2, T_2) = \sum_{k \in \{i \in \mathcal{N} \cup \{0\} : (t_{2i}(0), T_{2i}(0)) = (t_2, T_2)\}} r_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot))(k).$$

By hypothesis, therefore, we may write

$$\pi_2(\boldsymbol{\gamma}_2^0, \boldsymbol{\alpha}_2^0(\cdot), 0)(t_2, T_2) = \sum_{k \in \{i \in \mathcal{N} \cup \{0\} : (t_{2i}(0), T_{2i}(0)) = (t_2, T_2)\}} r_2(\boldsymbol{\gamma}_2^1, \boldsymbol{\alpha}_2^1(\cdot))(k).$$

But, by hypothesis for all $k \in \{i \in \mathcal{N} \cup \{0\} : (t_{2i}(0), T_{2i}(0)) = (t_2, T_2)\}$ it must be the case that

$$(t_{2k}(1), T_{2k}(1)) \in \Psi(t_2, T_2).$$

Thus, we must have that

$$\sum_{(t_2',T_2')\in\Psi(t_2,T_2)} \pi_2(\gamma_2^1, \boldsymbol{\alpha}_2^1(\cdot), 0)(t_2', T_2')$$

$$\geq \sum_{k\in\{i\in\mathcal{N}\cup\{0\}:(t_{2i}(0),T_{2i}(0))=(t_2,T_2)\}} r_2(\gamma_2^1, \boldsymbol{\alpha}_2^1(\cdot))(k).$$

This proves the second condition.

PROOF OF LEMMA 1:

The idea of the proof is to treat the fraction of middle class individuals in the population as a parameter and to show that as this fraction becomes very small: (i) the gains for the poor from making the investment go to zero while the costs remain positive, and (ii) the costs for the rich from making the investment go to zero while the benefits remain positive. The result will then follow from continuity.

Define $\lambda = \#\mathcal{N}_M/N$. Given our earlier assumptions, $\#\mathcal{N}_P/N = 1/2$ and $\#\mathcal{N}_R/N = 1/2 - \lambda$. We now define the variables of the example as functions of λ. Thus, $A(g, \lambda) = a_P^2/2 + \lambda(a_M + \delta g)^2 + (1/2 - \lambda)a_R^2$, $t_{2P}(g, \lambda) = [A(g, \lambda) - a_P^2]/[2A(g, \lambda) - a_P^2]$, $t_{2M}(g, \lambda) = \max\{0, [A(g, \lambda) - (a_M + \delta g)^2]/[2A(g, \lambda) - (a_M + \delta g)^2]\}$, and $\bar{y}(t, g, \lambda) = (1 - t)A(g, \lambda)$. Observe for future reference that $\lim_{\lambda\to 0} A(1, \lambda) = \lim_{\lambda\to 0} A(0, \lambda) = (a_R^2 + a_P^2)/2$, $\lim_{\lambda\to 0} t_{2P}(1, \lambda) = \lim_{\lambda\to 0} t_{2P}(0, \lambda) = [(a_R^2 - a_P^2)/2]/a_R^2$, and

$$\lim_{\lambda\to 0} t_{2M}(0, \lambda) = \frac{(a_R^2 + a_P^2)/2 - a_M^2}{a_R^2 + a_P^2 - a_M^2} > \lim_{\lambda\to 0} t_{2M}(1, \lambda)$$

$$= \max\left\{0, \frac{(a_R^2 + a_P^2)/2 - (a_M + \delta)^2}{a_R^2 + a_P^2 - (a_M + \delta)^2}\right\}.$$

Using the fact that $v(t, T, a) = T + [(1 - t)a]^2/2$, it is straightforward to calculate period-two expected utility for type J citizens ($J \in \{P, R\}$) as:

$$\tfrac{1}{2}[t_{2M}(g)\cdot\bar{y}(t_{2M}(g), g, \lambda) + (1 - t_{2M}(g))^2\cdot a_J^2/2]$$
$$+ \tfrac{1}{2}[t_{2P}(g)\cdot\bar{y}(t_{2P}(g), g, \lambda) + (1 - t_{2P}(g))^2\cdot a_J^2/2].$$

Thus, letting, $\omega_{JK}(g, \lambda) = t_{2K}(g, \lambda)\cdot\bar{y}(t_{2K}(g), g, \lambda) + (1 - t_{2K}(g, \lambda))^2\cdot a_J^2/2$, for $J \in \{P, R\}$ and $K \in \{P, M\}$, we know that a poor policy maker will not undertake the investment if $1/2[\omega_{PP}(1, \lambda) + \omega_{PM}(1, \lambda)]$ is smaller than $1/2[\omega_{PP}(0, \lambda) + \omega_{PM}(0, \lambda)]$, while a rich policy maker will undertake the investment if $1/2[\omega_{RP}(1, \lambda) + \omega_{RM}(1, \lambda)]$ is larger than $1/2[\omega_{RP}(0, \lambda) + \omega_{RM}(0, \lambda)]$.

Note that for $J \in \{P, R\}$ and $K \in \{P, M\}$

$$\omega_{JK}(0, \lambda) - \omega_{JK}(1, \lambda)$$
$$= (1 - t_{2K}(0, \lambda))[t_{2K}(0, \lambda)A(0, \lambda) + (1 - t_{2K}(0, \lambda))a_J^2/2]$$
$$- (1 - t_{2K}(1, \lambda))[t_{2K}(1, \lambda)A(1, \lambda) + (1 - t_{2K}(1, \lambda))a_J^2/2].$$

It is clear from our claims above that $\lim_{\lambda\to 0} \omega_{JP}(1, \lambda) = \lim_{\lambda\to 0} \omega_{JP}(0, \lambda)$, for $J \in \{P, R\}$. However,

$$\lim_{\lambda\to 0} [\omega_{JM}(0, \lambda) - \omega_{JM}(1, \lambda)]$$
$$= f_J\left(\lim_{\lambda\to 0} t_2^M(0, \lambda)\right) - f_J\left(\lim_{\lambda\to 0} t_2^M(1, \lambda)\right),$$

where $f_J(t) = (1 - t)[t(a_R^2 + a_P^2)/2 + (1 - t)a_J^2/2]$ for $J \in \{P, R\}$. Differentiating the function $f_J(t)$, we obtain $f_J'(t) = (a_{-J}^2 - a_J^2)/2 - ta_{-J}^2$, $J \in \{P, R\}$, where $-J = R$ when $J = P$ and $-J = P$ when $J = R$. Observe that $f_P'(t) > 0$ for all $t \leq (a_R^2 - a_P^2)/2a_R^2$. Thus, since $\lim_{\lambda\to 0} t_{2M}(1, \lambda) < \lim_{\lambda\to 0} t_{2M}(0, \lambda) < (a_R^2 - a_P^2)/2a_R^2$, we have that $\lim_{\lambda\to 0} \omega_{PM}(1, \lambda) < \lim_{\lambda\to 0} \omega_{PM}(0, \lambda)$. In addition, we have that $f_R'(t) < 0$ for all t which implies that $\lim_{\lambda\to 0} \omega_{RM}(1, \lambda) > \lim_{\lambda\to 0} \omega_{RM}(0, \lambda)$. By continuity, therefore, for sufficiently small λ, $1/2[\omega_{PP}(1, \lambda) + \omega_{PM}(1, \lambda)]$ is smaller than

$\frac{1}{2}[\omega_{PP}(0, \lambda) + \omega_{PM}(0, \lambda)]$, and $\frac{1}{2}[\omega_{RP}(1, \lambda) + \omega_{RM}(1, \lambda)]$ exceeds $\frac{1}{2}[\omega_{RP}(0, \lambda) + \omega_{RM}(0, \lambda)]$.

REFERENCES

Aghion, Philippe and Bolton, Patrick. "Government Debt and the Risk of Default: A Politico-Economic Model of the Strategic Role of Debt," in Rudiger Dornbusch and Mario Draghi, eds., *Public debt management: Theory and history*. Cambridge: Cambridge University Press, 1990, pp. 121–45.

Becker, Gary. "Public Policies, Pressure Groups, and Dead Weight Costs." *Journal of Public Economics*, December 1985, *28*(3), pp. 329–47.

Besley, Timothy and Coate, Stephen. "An Economic Model of Representative Democracy." *Quarterly Journal of Economics*, February 1997, *112*(1), pp. 85–114.

Brennan, Geoffrey and Buchanan, James M. *The power to tax: Analytical foundations of a fiscal constitution*. Cambridge: Cambridge University Press, 1980.

Buchanan, James, M. "The Relevance of Pareto Optimality." *Conflict Resolution*, December 1962, *6*(4), pp. 341–54.

———. "Toward Analysis of Closed Behavioral Systems," in James M. Buchanan and Robert Tollison, eds., *Theory of public choice*. Ann Arbor, MI: University of Michigan Press, 1972, pp. 11–23.

Coase, Ronald. "The Problem of Social Cost." *Journal of Law and Economics*, October 1960, *3*, pp. 1–44.

Coate, Stephen and Morris, Stephen. "On the Form of Transfers to Special Interests." *Journal of Political Economy*, December 1995, *103*(6), pp. 1210–35.

Dixit, Avinash. *The making of economic policy: A transaction-cost politics perspective*. Cambridge, MA: MIT Press, 1996.

Dixit, Avinash and Londregan, John. "Redistributive Politics and Economic Efficiency." *American Political Science Review*, December 1995, *89*(4), pp. 856–66.

Glazer, Amihai. "Politics and the Choice [illegible] Durability." *American Economic Review*, December 1989, *79*(5), pp. 1207–13.

Krusell, Per; Quadrini, Vincenzo and Rios-Rull, Jose-Victor. "Politico-Economic Equilibrium and Economic Growth." *Journal of Economic Dynamics and Control*, January 1997, *21*(1), pp. 243–72.

Meltzer, Allan H. and Richard, Scott F. "A Rational Theory of the Size of Government." *Journal of Political Economy*, October 1981, *89*(5), pp. 914–27.

Milesi-Ferretti, Gian-Maria and Spolaore, Enrico. "How Cynical Can an Incumbent Be? Strategic Policy in a Model of Government Spending." *Journal of Public Economics*, September 1994, *55*(1), pp. 121–40.

Osborne, Martin J. and Slivinski, Al. "A Model of Political Competition with Citizen Candidates." *Quarterly Journal of Economics*, February 1996, *111*(1), pp. 65–[illegible].

Persson, Torsten and Svensson, Lars. "Why a Stubborn Conservative Would Run A Deficit: Policy with Time-Inconsistent Preferences." *Quarterly Journal of Economics*, May 1989, *104*(2), pp. 325–46.

Rogoff, Kenneth. "Equilibrium Political Budget Cycles." *American Economic Review*, March 1990, *80*(1), pp. 21–36.

Romer, Thomas. "Individual Welfare, Majority Voting and the Properties of a Linear Income Tax." *Journal of Public Economics*, July 1975, *4*(2), pp. 163–85.

Stigler, George. "Economists and Public Policy." *Regulation*, May/June 1982, pp. [illegible]–13.

Tabellini, Guido and Alesina, Alesina. "Voting on the Budget Deficit." *American Economic Review*, March 1990, *80*(1), pp. [illegible]–49.

Williamson, Oliver. "The Politics and Economics of Redistribution and Efficiency." Mimeo, University of California-Berkeley, 1994.

Wittman, Donald. "Why Democracies Produce Efficient Results." *Journal of Political Economy*, December 1989, *97*(6), pp. 1395–426.

Altruists, Egoists, and Hooligans in a Local Interaction Model

By Ilan Eshel, Larry Samuelson, and Avner Shaked*

We study a population of agents, each of whom can be an Altruist or an Egoist. Altruism is a strictly dominated strategy. Agents choose their actions by imitating others who earn high payoffs. Interactions between agents are local, so that each agent affects (and is affected by) only his neighbors. Altruists can survive in such a world if they are grouped together, so that the benefits of altruism are enjoyed primarily by other Altruists, who then earn relatively high payoffs and are imitated. Altruists continue to survive in the presence of mutations that continually introduce Egoists into the population. (*JEL* C70, C78)

An act is altruistic if it confers a benefit on someone else while imposing a cost on its perpetrator. How does costly altruistic behavior survive? Why doesn't utility maximization inexorably eliminate such behavior?

One answer is immediately available: allegedly altruistic acts are not really altruistic. Upon closer examination, they confer net benefits rather than costs. For example, charitable donations may bring benefits such as public recognition or a warm glow that overwhelm the cost of the donation. If we push revealed preference theory to its logical limit, this conclusion becomes as inescapable as it is tautological. If someone commits an "altruistic" act, then this reveals that he prefers doing so.[1]

A second answer is also available: the interaction in which the altruistic act occurs may be repeated. If the interaction is infinitely repeated, then the folk theorem (Drew Fudenberg and Eric Maskin, 1986) ensures that there are equilibria in which Altruists survive, though there are also equilibria in which altruism does not appear.[2] David M. Kreps et al. (1982) show that there are equilibria in which Altruists survive in finitely repeated games with incomplete information, though once again a folk theorem result appears, including equilibria without altruism (Fudenberg and Maskin, 1986).

We do not doubt that people often derive benefits from seemingly altruistic acts, and that many interactions are repeated. However, we also believe that altruistic acts occur for which conventional models do not readily account. Embellishing the models to encompass such acts often leads to utility functions that are uncomfortably exotic or to an uncomfortably strong faith in repetition.

This paper provides an alternative model of altruistic behavior with two key properties. First, we abandon the assumption that people

* Eshel: Department of Statistics, School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel; Samuelson: Department of Economics, University of Wisconsin, 1180 Observatory Drive, Madison, WI 53706; Shaked: Department of Economics, University of Bonn, Adenauerallee 24-26, D 53113 Bonn, Germany. We thank Ed Green, Reinhard Selten, and an anonymous referee for helpful comments. Financial support from the National Science Foundation and the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303 at the University of Bonn, is gratefully acknowledged. Part of this work was done while Eshel and Samuelson were visiting the Department of Economics at the University of Bonn and while Samuelson and Shaked were visiting the Institute for Advanced Studies at the Hebrew University of Jerusalem. We are grateful to both for their hospitality and support.

[1] There are many parallels between discussions of altruism in economics and biology. Biologists frequently argue that once the full effects of actions on reproductive success are recognized, seemingly altruistic acts are privately optimal. For example, there are now numerous explanations for why it may be individually beneficial for a bird to attract a predator's attention by uttering a cry that alerts the bird's flock to the predator's presence. Richard Dawkins (1976) discusses bird alarms as well as other cases of apparent altruism.

[2] In biological contexts, explanations for altruism similar to those that arise in repeated games appear in the guise of reciprocal altruism (Robert L. Trivers, 1971).

are rational agents choosing utility-maximizing actions. Instead, we believe that people must learn which actions work well, and that an important force in learning is imitation. Second, interactions between agents in our model are "local," meaning that altruistic acts are more likely to affect nearby agents than more distant neighbors and that agents are more likely to imitate nearby than more distant neighbors.

To see how these forces can allow altruism to survive, suppose there are two kinds of agents: Altruists, who provide a public good to their neighbors at a cost to themselves, and Egoists, who do not do so. Suppose further that Altruists tend to exist in concentrated groups. Altruists can then earn higher payoffs than Egoists, because Altruists are more likely to enjoy the public goods provided by other Altruists. The imitation-based learning process now prompts other agents to become Altruists. In addition, nearby agents are the ones most likely to imitate the Altruists. This preserves the tendency of Altruists to clump together in groups and hence preserves the conditions needed for altruism to survive.

This argument is unconvincing without some stability analysis. We expect perturbations to occasionally switch the behavior of some agents from Altruist to Egoist or Egoist to Altruist, perhaps because someone has analyzed the model and deduced that it is utility maximizing to be an Egoist, or has made a mistake, or has simply experimented with a new action. An Egoist thrust into the midst of Altruists will thrive on the public goods provided by the latter and will be imitated, while an Altruist thrust in the midst of Egoists will fare poorly and will be ignored. Perturbations or "mutations" that occasionally cause people to switch strategies thus apparently produce a force pushing toward egoism. An explanation of altruistic behavior must demonstrate that altruism can withstand such mutations.[3]

[3] Group selection models share many of the features of our model of local interaction, but are now widely considered to be implausible as explanations for altruistic behavior, largely because of their inability to withstand mutations. We discuss the relationship of our work to group selection arguments in Section V.

We find that only states composed prim of Altruists survive in the presence of rare tations. If a mutation introduces an Egoist the midst of Altruists, then the Egoist will vive and spread. However, the resulting gr of Egoists quickly confronts limits on its a ity to expand, as each expansion causes public goods supplied by neighboring Al ists to be shared among more and more Ego and hence reduces Egoists' payoffs. Ego are thus readily introduced but cannot exp beyond small, isolated groups. Isolated Al ists, in contrast, cannot even survive in midst of Egoists. However, mutations will casionally introduce a group of Altruists in midst of Egoists. Such a group of Altruists expand without bound. Mutations thus m readily lead to large groups of Altruists Egoists, allowing the former to dominate.

Altruism is not the only type of extern that can arise between agents. We extend model to consider Hooligans, or agents wh benefit from imposing damages on th neighbors. The same forces that allow Al ists to survive in the presence of Egoists allow Altruists to survive against Hoolig or Egoists to survive in the midst of Hoo gans, though Hooligans will typically not eliminated entirely. The analysis is then ther extended to general 2×2 games, all ing us to examine games with two strict N equilibria, one payoff dominant and one dominant.

The work most closely related to ours cludes Theodore C. Bergstrom and Oded St (1993), Lawrence E. Blume (1993), Glenn Ellison (1993), and papers by M A. Nowak and Robert M. May (1992, 19 and Nowak et al. (1994). Our spatial stru matches one of the models considered Bergstrom and Stark as well as the sim case considered by Ellison, while Blume Nowak and May examine spatial mode which agents are arranged in a plane than along a line. We differ from Elliso Blume in taking imitation, rather than variant of best-reply dynamics, to be the ing force behind strategy selections. Th crucial, as altruism has no hope in a wor best responders. We differ from Nowak May in relying on analytical techniques than simulations, albeit for a very si

model, and especially in studying the effect of mutations.

Section I presents the model. Section II examines equilibria of the imitation process. There are many possible limiting outcomes, depending upon the initial conditions of the system, but Altruists comprise a significant portion of the population in all but one of these. We establish conditions under which the probability of an initial condition leading to the elimination of altruism shrinks to zero as the population grows large. Section III shows that only those limiting outcomes with a significant proportion of Altruists survive mutations. Section IV pursues a generalization of the model in which agents interact in larger neighborhoods. In doing so, we find that it can be to Altruists' advantage to have a relatively high cost of altruism. A higher cost of altruism ensures that if Altruists survive, then they must do so in larger groups, because only then do they share enough of the public good to compensate for the high cost of being an Altruist. This in turn ensures that if there are any Altruists at all, then there are relatively large groups of Altruists. Section V concludes. Unless otherwise noted, proofs are contained in the Appendix.

I. Altruists and Egoists

We consider a collection of N individuals, where N is finite. Each individual can be either an Altruist or an Egoist. An Altruist provides a public good that contributes one unit of utility to those who receive its benefits. The *net* cost to the Altruist of providing the public good is $C > 0$, so that the combination of enjoying the benefits of his own public good and bearing the costs of its provision reduces the Altruist's utility by C. Egoists provide no public goods and bear no costs. Instead they simply enjoy the benefits of the public goods provided by others.

Time is divided into discrete periods. At the end of each period, after consuming any public good that is available and bearing provision costs (if an Altruist), each agent decides, according to a learning rule, whether to be an Altruist or Egoist in the next period.

The nature of this learning rule is important. One possibility is that the agents are fully rational, though myopic, and the learning rule leads them to adopt single-period, expected-utility-maximizing actions. In this case, they will realize they face a variant of the prisoner's dilemma and will play the strictly dominant strategy, namely Egoist. Instead of choosing best replies, however, our players imitate the strategies of others whom they observe to be earning high payoffs.

At one level, this imitation seems preposterous. How hard can it be to figure out that being an Egoist (or defecting in a prisoner's dilemma) is a strictly dominant strategy? This is indeed a trivial task for a game theorist facing the sterilized 2×2 games with which we often work. However, these games are a simplified representation of a much more complicated reality. The agents who actually play the game may not recognize that they are playing a game, may not know who their opponents are, may not know what strategies are available, and may not know what payoffs these strategies bring. They may then be unable to think like game theorists, or like the agents in game-theoretic models. At the same time, we believe that people are generally able to form a good estimate of others' payoffs, whether these payoffs are measured in terms of money or other units such as social status or prestige, and that people tend to imitate the behavior of those they observe earning high payoffs.

Imitation alone appears to hold out no hope for the survival of altruism. Egoists will enjoy the same public goods as Altruists, while only the latter bear costs. As a result, all Egoists will earn higher payoffs than all Altruists and imitation can only lead players to become Egoists.

This argument is applicable only if the benefits of the public good provided by each Altruist extend to every agent in the population. The prospects for Altruists improve if the public good is a *local* public good. To make this precise, we introduce a neighborhood structure taken from Bergstrom and Stark (1993) and Ellison (1993). Agents in the model are located around a circle.[4] Each agent interacts

[4] A spatial interpretation is convenient, but local interaction structures may arise in other ways. In academia, field of specialization is probably more important than location in determining patterns of interaction.

with his two immediate neighbors, i.e., with one agent to his right and one to his left. If an agent is an Altruist, then his immediate neighbors enjoy the benefit of his public good provision. The payoff of agent i is then given by $N_i^A - C$ if i is an Altruist and N_i^A if i is an Egoist, where $N_i^A \in \{0, 1, 2\}$ is the number of i's Altruist neighbors (excluding himself).

In each period, each agent takes a draw from an independent Bernoulli trial, causing the agent to "learn" with probability $\mu \in (0, 1]$ and to retain her strategy with probability $1 - \mu$. An agent who learns observes her own payoff and the payoff and strategy of each agent in her neighborhood. She then chooses to be an Egoist if the average payoff of the Egoists in her sample exceeds that of Altruists, and chooses to be an Altruist if the average payoff of Altruists exceeds that of Egoists.[5] If an agent and her two neighbors all play the same strategy, be it Altruist or Egoist, then the agent will continue to play that strategy.

We shall concentrate on the case of $\mu = 1$, so that every agent learns in every period. This gives a deterministic learning process which simplifies the derivation and statement of the results. However, Bernardo A. Huberman and Natalie S. Glance (1993) have recently argued that the outcomes of local interaction models can be sensitive to whether all agents adjust their strategies at the same time. We accordingly comment on how each of our results would be modified if $\mu < 1$.

A state is a specification of which agents are Altruists and which are Egoists. Let S be the set of possible states. For states i and j in S, let P_{ij} be the probability that a single iteration of the imitation process changes the system to the state j given that the current state is i. Since the learning process is deterministic (with $\mu = 1$), P_{ij} is either 0 or 1. The collection $\{P_{ij}\}_{i,j \in S}$, along with a specification of the initial state at time zero, is a Markov process on the state space S. We refer to this Markov process as the "imitation dynamics."

II. Equilibrium

A. *Absorbing Sets*

We are interested in the stationary distributions of the imitation dynamics. We say that a set of states is *absorbing* if it is a minimal set of states with the property that the Markov process can lead into this set but not out of it. An absorbing set may contain only one state, say i, in which case $P_{ii} = 1$ and i is a stationary state of the Markov process. An absorbing set may contain more than one state, in which case $P_{ij} = 0$ if i is contained in the absorbing set and j is not, while the Markov process cycles between states in the absorbing set.

For each absorbing set of the Markov process, there is a unique stationary distribution the support of which consists of that absorbing set. We can then learn much about the stationary distribution of the learning process by studying absorbing sets.

We begin by compiling a description of the imitation dynamics. We assume $C < 1/2$.[6] At the end of each period, an agent may either retain her strategy or choose a strategy played by one of the two agents closest to her, depending upon their payoffs. These payoffs in turn depend on the strategies of the next two neighbors. The fate of an individual is thus completely determined by the strategies of her four nearest neighbors.

An Egoist who learns by imitating his neighbors can become an Altruist only if at least one of his two nearest neighbors is an Altruist. However, if both of his immediate neighbors are Altruists, then the Egoist earns

[5] We simplify the analysis by choosing the cost C so that payoff ties do not arise. There are many other plausible learning rules. For example, an agent may simply compare the best Egoist and best Altruist payoff among those payoffs she observes, or may compare the sum of the Egoist and Altruist payoffs, rather than considering averages. Itzhak Gilboa and David Schmeidler (1995, 1996), in the context of their case-based decision theory, examine the difference between considering the sums or the averages of payoffs.

[6] We can always vanquish altruistic behavior by making it too expensive. If $C > 1/2$, we find that the only absorbing sets of the learning process are those consisting of single states in which either all agents are Altruists or all are Egoists. The former contains only itself in its basin of attraction, while the latter attracts the remainder of the state space. In the presence of mutations, only the latter absorbing state survives. The case of $C = 1/2$ leads to inconvenient payoff ties.

a payoff of two, more than an Altruist can ever earn, causing the Egoist to retain his strategy. An Egoist can therefore become an Altruist only if exactly one of his neighbors is an Altruist, with the Altruist neighbor earning a higher payoff than the average payoffs of the Egoist and his Egoist neighbor. This payoff inequality holds only if the Altruist has an Altruist neighbor, since otherwise the Altruist receives the lowest possible payoff of $-C$, and if the other Egoist in the neighborhood faces a neighborhood containing only Egoists, so as to bring the average Egoist payoff below $1 - C$. Hence, an Egoist can become an Altruist only if he faces either the following combination of strategies or its mirror image, where "a" represents an Altruist and "E" an Egoist,

(1) $\qquad aa \quad E \quad EE$

where it is the central Egoist who converts to an Altruist.[7] In all other cases, Egoists remain Egoists.

A similar calculation shows that an Altruist will remain an Altruist if and only if one of the following combinations of strategies (or their mirror images) occurs,

(2) $\qquad xa \quad a \quad ax$

$\qquad\quad aa \quad a \quad EE$

where it is the central Altruist whose fate is in question and where an "x" holds the place of an agent who may be either an Altruist or an Egoist. In all other cases, Altruists change to Egoists.

Conditions (1) – (2) provide a complete description of the individual imitation dynamics. To illustrate some absorbing sets, we represent the agents as being located on a line, where we think of the ends of the line as being joined to form a circle. From (1) – (2), we easily verify that the following are absorbing sets:

- The state in which all are Altruists.
- The state in which all are Egoists.
- A state in which all are Altruists except two adjacent Egoists:

$$\cdots aaaaaaaaEEaaaaaaaa \cdots .$$

- A set of two states, consisting of:

$$\cdots aaaaaaaaaEaaaaaaaaa \cdots$$

$$\cdots aaaaaaaaEEEaaaaaaaa \cdots .$$

In this last case, the imitation dynamics cycle between the two states in the absorbing set. The lone Egoist initially earns the highest possible payoff of 2, inducing his two neighbors to become Egoists and leading to the second state in the cycle. Each of these new Egoists finds himself in the situation described by (1), where he has two Egoists on one side and two Altruists on the other. This causes the new Egoists to switch back to altruism, beginning the cycle anew. We refer to such a cycle as a *blinker*.

The two outside agents in the blinker face a coordination problem. It is an equilibrium for one but not for both to be an Egoist, and the learning scheme causes them to cycle around this equilibrium. We suspect that cycles in behavior do occur, though our simple model captures these cycles in a crude way. The presence of blinkers is a product of setting $\mu = 1$, forcing all agents to assess their strategies in every period. If $\mu < 1$, then blinkers are no longer absorbing sets, since a period will eventually arise in which only one of the two outside agents in the blinker revises her strategy, leading to a pair of adjacent Egoists. All absorbing sets would then be singletons.

These examples, and combinations constructed from them, include all of the possibilities for absorbing sets. Some terms will be useful in making this precise. If agents α and β play the same strategy, either Altruist or Egoist, and if all agents between α and β play this strategy, then we will refer to agents α, β, and the intermediate agents as an interval of

[7] We find the displays easiest to read if we use a lower case "a" to represent Altruists, and the text easiest to read if we continue to use "A." We will also often separate agents in whom we are interested by spaces, as in the case of the central Egoist here, though these spaces have no significance other than directing attention to particular

either Altruists or Egoists. We call a maximal such interval a *string*. Notice that strings may be of any length from 1 to N, the length of the circle. We then have (the proof is in the Appendix):[8]

PROPOSITION 1: *Let* $0 < C < \frac{1}{2}$ *and* $\mu = 1$. *Then:*

(1.1) *Absorbing sets consist of (i) the state in which all agents are Egoists, (ii) the state in which all agents are Altruists, and (iii) sets containing states in each of which Altruist strings of length three or longer are separated by Egoist strings of length less than four. These sets are either singletons (in which case all Egoist strings are of length two) or contain two states [in which case any string of length one (three) in one of the states blinks to a string of length three (one) in the other].*

(1.2) *Except for the state in which all agents are Egoists, the proportion of Altruists in an absorbing state, or the average proportion over the two states in an absorbing set, is at least* 0.6.

Proposition 1 indicates that there are many absorbing sets, each of which is the support of a stationary distribution of the imitation process. In all but one of these absorbing sets, the majority of the population is Altruists. Hence, there is no possibility for moderation in altruism. If Altruists survive at all, they must be the majority.

To see what lies behind this result, we first note that a string of Egoists in an absorbing set can never be longer than three. If the length of an Egoist string exceeds three, then the two Egoists at its edges will each have two Egoists on one side and two Altruists on the other, and hence they will become Altruists [cf. (1)], causing the string to shrink. Egoists can thus survive only in strings of length two or strings of length one (where the latter alternate with strings of length three in a blinker). Altruist strings must be at least length three in order to survive, and surviving Altruist strings can expand, since doing so creates more and more high-payoff Altruists. This allows us to conclude that if there are any Altruists at all, then Altruists will occur in strings of length at least three while Egoists occur in strings of at most two (or in blinkers the average length of which is two), and hence there will be at least 60-percent Altruists.

B. *Basins of Attraction*

Because the state in which all agents are Egoists is absorbing, the system may drive Altruists to extinction. To assess the likelihood of such an event, we identify the initial conditions from which the system converges to an absorbing set containing Altruists.

The proof of Proposition 1 shows that a string of Altruists either drops below length three at some point, after which it disappears, or persists forever. We refer to a string of Altruists whose fate is the latter as a "persistent" string. The system will converge to a state in which at least 60 percent of the agents are Altruists if and only if the initial condition contains at least one persistent string.

The following proposition first characterizes persistent strings. We then suppose the agents' initial identities as either Altruists or Egoists are randomly determined, and investigate the probability that this leads to an initial state containing a persistent string of Altruists. We have:[9]

PROPOSITION 2: *Let* $0 < C < \frac{1}{2}$ *and* $\mu = 1$. *Then:*

(2.1) *A string of Altruists is persistent if and only if (i) the string contains at least five Altruists, (ii) the string consists of four Altruists bordered on at least one end by two Egoists, or (iii) the string consists of three Altruists bordered on each end by two Egoists or bordered on at least one end by three Egoists. All other strings of Altruists are eliminated by period three.*

(2.2) *If agents' initial identities as Altruists or Egoists are determined by independent identically distributed random variables*

[8] If $\mu < 1$, then this proposition continues to hold, except that blinkers are no longer absorbing sets.

[9] This result holds if $\mu < 1$, with the modification that some strings of two Altruists, as well as strings of the form aEaaaEEE, survive with a probability greater than zero but less than one.

has positive probability on Altruist, then as N gets large, the probability of a persistent string of Altruists in the initial state, and hence convergence to an absorbing set containing at least 60-percent Altruists, approaches unity.

Under randomly determined initial conditions and a large population, the probability that Altruists survive is high because there will almost certainly be an initial group of Altruists large enough to ensure their survival, and hence to ensure that most agents are eventually Altruists. However, a great deal of growth may be required before a single group of Altruists can comprise an appreciable fraction of a large population. How long must we wait before most agents are Altruists?

By period three, any string of Altruists that is not persistent will have been eliminated, and the population will consist of persistent strings of Altruists separated by strings of Egoists. If a string of Egoists is not already of length two or a blinker, then it will contract at a rate of two agents per period, as the Egoists on the two ends of the string switch to altruism. We can accordingly pose our waiting-time question as the following: how long do we expect to wait until a string of agents lying between two persistent Altruist strings has been reduced to length two or to a blinker? But since this waiting time is half of the string's length, plus possibly three periods, we can equivalently ask how long a string of agents we expect to find between two persistent strings of Altruists. There is no reason why strings of Egoists lying between surviving strings of Altruists should be longer in larger populations, and hence no reason for expected waiting times to increase as N increases. However, the expected length of such a string is quite sensitive to the probability that an agent in the initial condition is an Altruist. We can calculate:[10]

PROPOSITION 3: *Let $0 < C < 1/2$ and $\mu = 1$, and let agents' identities in the initial condition be randomly and independently determined, with probability p attached to being an Altruist. Then, in the limit as the population gets arbitrarily large, an upper bound on the expected length of a string of agents between two persistent Altruist strings is given by:*

p	Length
0.5	22
0.4	34
0.3	63
0.2	171
0.1	1140
0.05	8480
0.01	1.01×10^6
0.001	1.00×10^9.

Expected waiting times are thus relatively moderate as long as there is a sufficiently high initial probability that a randomly selected agent is an Altruist. For example, a probability of altruism of 10 percent gives an upper bound of 573 on the expected waiting time (half of the Egoist string's expected length, plus three). On the other hand, persistent strings will be extremely rare if Altruists are very rare, and expected waiting times will be very long.[11]

C. *Hooligans*

Altruism, conferring a benefit on someone else at a cost to oneself, is not the only way that one agent's actions may affect another. At the opposite extreme we have Hooligans, who benefit by imposing harm on others. Notice that hooliganism need not be limited to the psychopathic. Those who litter in order to avoid the cost of disposing of their refuse,

[10] If $\mu < 1$, then agents learn less frequently, and expected waiting times will be longer.

[11] We expect to wait longer until Altruists dominate the population when Altruists are rare, but the resulting absorbing states are likely to have higher proportions of Altruists. This result holds because such initial conditions will be characterized by relatively small numbers of long strings of Egoists, who will be transformed into small numbers of short strings of Egoists. When Altruists are more likely, there will be many short but distinct strings of Egoists in the initial condition, leading to an absorbing state with more strings of Egoists.

		II	
		X	Y
I	X	a, a	b, c
	Y	c, b	d, d

FIGURE 1. 2×2 GAME

those who pollute rather than take costly abatement measures, and those who shirk in group efforts are all Hooligans.

Our model is easily generalized to accommodate Hooligans. Let there be two types of agents, denoted by 1 and 2. Let type 1 contribute K_1 to the payoff of each of his neighbors at a cost of C_1 to himself. Let type 2 contribute K_2 to each neighbor at a cost C_2 to himself. There is no loss of generality in assuming that $K_1 > K_2$. Our model of Altruists and Egoists is then the special case in which $K_1 = 1$, $C_1 = C$, and $K_2 = C_2 = 0$. The behavior of the model depends only on a single parameter:[12]

PROPOSITION 4: *Let $K_1 > K_2$. Then any variation in the values of K_1, K_2, C_1, and C_2 that preserves*

$$\frac{C_1 - C_2}{K_1 - K_2} \tag{3}$$

gives rise to the same imitation dynamics.

Hence, any two specifications of the payoffs that preserve $(C_1 - C_2)/(K_1 - K_2)$ give rise to the same absorbing sets, basins of attraction, and dynamic paths for the imitation dynamics, and the same limiting distributions in the presence of mutations.

For the Altruist and Egoist model of the previous sections, the ratio (3) was C, which was interpreted as the cost of altruism. Consider the following pairs of types of players. In each case, the first column identifies the effect an agent of type 1 has on his two neighbors and the cost to the agent of that effect, while the second column provides analogous information for an agent of type 2.

$$\begin{pmatrix} K_1 & K_2 \\ C_1 & C_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ C & 0 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 0 & -C \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ C & 0 \end{pmatrix},$$

$$\begin{pmatrix} 1 & -1 \\ C & -C \end{pmatrix}, \begin{pmatrix} -1 & -2 \\ -C & -2C \end{pmatrix}.$$

The first specification is the familiar Altruist and Egoist pair from previous sections. The second pair of agents consists of an Egoist and a Hooligan who enjoys (incurs a negative cost from) causing damage of one unit to his neighbors. In case this Hooligan seems too malicious in his enjoyment of the harm he causes, the third pair rewrites this situation as an agent of type 1 who imposes no harm on others but incurs a cost of C to avoid doing so, with a type-2 agent who does not incur the cost and imposes damage of one unit on his neighbors. The fourth pair includes an Altruist and a Hooligan. The last pair has two Hooligans, one of whom causes twice the damage and doubly benefits from doing so. In each of these specifications, the ratio $(C_1 - C_2)/(K_1 - K_2)$ is given by C, and hence these are equivalent models. As long as $C < 1/2$, it is always the first type in each pair that will come to comprise the majority of a large population with a randomly determined initial condition. Hooligans will then be in the minority when facing Egoists or Altruists, though some Hooligans will survive, just as some Egoists survive when paired against Altruists.[13]

Similar insights can be used to extend the analysis to general 2×2 symmetric games. Suppose each agent must choose a single strategy to use when playing the game shown in Figure 1 with each of his neighbors. With

[12] The proof establishes this proposition for a wide class of imitation dynamics including those of the current section as a special case.

[13] The final specification shows that Hooligans can be in the majority if paired with even worse Hooligans.

sacrificing generality, we can assume $a > d$. We will then further concentrate on the case in which $a > b$. This latter assumption excludes some games but retains the common examples of 2×2 games. Then an argument analogous to the proof of Proposition 4 shows that the imitation dynamics depends only upon the two numbers:

$$\alpha = \frac{c-b}{a-b}, \quad \beta = \frac{d-b}{a-b}. \tag{4}$$

In light of this, we can transform the payoffs in Figure 1 by subtracting b from each payoff and dividing by $a - b$ to obtain the equivalent representation of the game given in Figure 2, where $\alpha = (c - b)/(a - b)$ and $\beta = (d - b)/(a - b)$.

We can now classify games according to the values of α and β, where $\beta < 1$ (because we have assumed $a > d$). We have:

- Prisoner's Dilemma: $0 < \beta < 1$, $1 < \alpha$.
- Coordination Game: $0 < \beta < 1$, $\alpha < 1$.
- Chicken: $\beta < 0$, $1 < \alpha$.
- Efficient Dominant Strategy: $\beta < 0$, $\alpha < 1$.

This classification is illustrated in Figure 3. An "efficient dominant strategy" game is one in which X is a strictly dominant strategy and the outcome (X, X) is efficient, unlike the prisoner's dilemma. A coordination game has two strict Nash equilibria, given by (X, X) and (Y, Y). Chicken has one mixed-strategy Nash equilibrium and two asymmetric pure strategy equilibria.[14] In the case of a coordination game, (X, X) is the payoff-dominant equilibrium (because $a > d$ and hence $\beta < 1$), and is also risk dominant if $\alpha + \beta < 1$, while the equilibrium (Y, Y) is risk dominant if $\alpha + \beta > 1$. The interval in which $\alpha = 1 + \frac{1}{2}C$, $\beta = \frac{1}{2}C$, and $C < \frac{1}{2}$, shown in Figure 3, describes the range of Altruist and Egoist games that was analyzed in subsections A and B of this section.

		II: X	II: Y
I	X	$1, 1$	$0, \alpha$
	Y	$\alpha, 0$	β, β

FIGURE 2. TRANSFORMATION OF GAME IN FIGURE 1

The methods developed in the previous sections to examine Altruists and Egoists can be applied to any other game in this classification. For example, consider coordination games. Let $\alpha + \beta > 1$ so that (X, X) is the payoff-dominant equilibrium but (Y, Y) is the risk-dominant equilibrium. Now consider a boundary between a group of agents playing strategy X and a group playing strategy Y, or[15]

$$\cdots xxxxxxxxxxxYYYYYYYYYYYY \cdots.$$

The only agents at risk of changing their strategies are the two agents, one playing X and one playing Y, at the ends of their respective strings. Each faces a neighborhood with one X and one Y agent, in addition to themselves. In Ellison's (1993) model, each chooses a best response to his two neighbors. By assumption, Y is risk dominant and hence is a best reply when one neighbor plays X and one plays Y. Hence, the agent playing Y retains his strategy while the agent playing X switches to Y. The string of Y's thus grows while the string of X's shrinks, ensuring that best-reply learning leads to the selection of the risk-dominant equilibrium.

In our imitation model, the X player on the boundary earns a payoff of 1, while the adjacent X player earns 2. The Y player on the boundary earns $\alpha + \beta$ while the adjacent Y player earns 2β. Comparing the average payoffs, we find that the boundary player Y retains his strategy if $\alpha + 3\beta > 2$, while a boundary

[14] The asymmetric pure strategy equilibria of this symmetric game become relevant if agents can condition their strategies on some asymmetry, such as location.

[15] As with Altruists and Egoists, we find the displays easier to read if we use a lower case x to represent the strategy X.

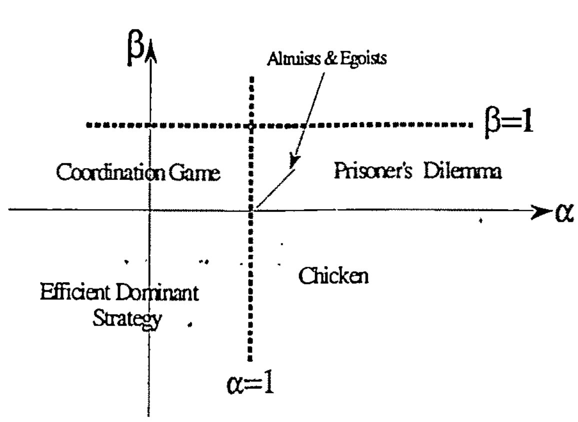


FIGURE 3. CLASSIFICATION OF GAMES

X will turn into Y if $\alpha + \beta > 3/2$. Hence, there are three subregions of (α, β) values within the region of coordination games in which risk dominance and payoff dominance conflict. In the first region, $\alpha + 3\beta > 2$ and $\alpha + \beta > 3/2$, and hence both players will play Y in the following period. In the second region, $\alpha + 3\beta < 2$ and $\alpha + \beta < 3/2$, and hence both will play X in the following period. In the third region, $\alpha + 3\beta > 2$ and $\alpha + \beta < 3/2$, and both will retain their strategy. In the first region, the string playing the risk-dominant action will grow, in the second it will shrink while the string playing the payoff-dominant action grows, and in the third region each string maintains its length.

If a string of agents playing the risk-dominant action Y is to expand, its payoffs must provide a premium over that required for risk dominance (i.e., $\alpha + \beta > 3/2$). This is necessary because an agent at the end of a string of X agents compares not whether X or Y is a best reply, but whether the X or Y players in his neighborhood are earning higher average payoffs. One of the X players in his neighborhood is bordered by two other X players, and hence receives an exceptionally high payoff. Risk dominance alone is not enough to overcome this payoff.

If a string of agents playing the *payoff-dominant* action X is to expand (while Y is risk dominant), then we must have $\alpha + 3\beta < 2$ and $\alpha + \beta < 3/2$. In conjunction with the requirement that $\alpha + \beta > 1$, these inequalities require $\alpha > \beta$. Hence, the payoff to playing strategy Y must be greatest if the opponent plays X, even though (Y, Y) is an equilibrium. This occurs because the neighborhood of an agent who plays Y, and who is located on the end of a string of Y agents, contains a Y player who faces two Y opponents and hence earns a relatively high payoff. The average payoff to X can be highest only if the payoff in equilibrium (Y, Y) is relatively small. This is in turn compatible with risk dominance only if $\alpha > \beta$.

Given that strategy Y must receive a premium over risk dominance in order to expand, and given that strategy X can expand in the absence of this premium only if the additional condition $\alpha > \beta$ holds, then it is no surprise that there are some cases in which neither string will expand. Imitation can then yield peaceful coexistence of the two strategies, unlike best-response behavior. Imitation allows the coordination failures created by coexistence to persist because agents on the boundary of a string, and hence experiencing coordination failures, are most likely to observe other agents with the same strategy who are not facing coordination failures and to observe agents with the other strategy who *are* plagued by such failures. This introduces a force against changing strategies, and builds sufficient inertia into the system to support coexistence.

III. Mutations

We now ask how altruism fares in the presence of mutations. We assume that at the end of each period, and after imitation has occurred, each agent takes a draw from an independent, identically distributed Bernoulli random variable. With probability λ, this agent is a mutant and changes his type, either from Altruist to Egoist or from Egoist to Altruist. With probability $1 - \lambda$, this agent experiences no mutation. We will be interested in the case in which λ is small, so that imitation is the primary force driving strategy revisions. We study this by examining the limiting case as the mutation probability λ goes to zero.[16]

[16] In economics, the common practice is to follow the lead of Michihiro Kandori et al. (1993) and H. P

Let Γ_{ij} be the probability that the combination of imitation and mutation changes the state to j given that the current state is i.[17] Then $\{\Gamma_{ij}\}_{i,j \in S}$ is again a Markov process on the state space S, which we refer to as the "imitation-and-mutation dynamics." Notice that $\Gamma_{ij} > 0$ for all i and j, which is to say that for any two states i and j, there is some combination of mutations capable of changing the system from i to j. Hence, for each fixed mutation rate, the imitation-and-mutation dynamics has a unique stationary distribution. The proportions of states reached along any sample path approach this distribution almost surely, and the distribution of states at time t approaches this distribution as t gets large. (John G. Kemeny and J. Laurie Snell [1960 Theorems 4.1.4, 4.1.6, and 4.2.1].)

We study the limit of these stationary distributions as the probability of a mutation λ gets small, which we refer to as the *limiting distribution*.

PROPOSITION 5: *Let* $0 < C < \frac{1}{2}$. *If* $N > 30$, *then the limiting distribution places positive probability only on states contained in absorbing sets of the imitation process in which the proportion of Altruists is at least* 0.6.

The techniques involved in establishing this result, which holds for $\mu \in (0, 1]$, were developed by M. I. Freidlin and A. D. Wentzell (1984) and were introduced into economics by Kandori et al. (1993) and Young (1993). The argument begins by observing that when the mutation rate is small, the system spends virtually all of its time in absorbing sets of the imitation dynamics, and hence the limiting distribution allocates all of its probability to such sets. Movements between absorbing sets of the imitation dynamics can be accomplished only by mutations. The system will allocate most of its probability to absorbing sets of the imitation process that are easy to reach, in the sense that it requires relatively few mutations to reach their basin of attraction from other absorbing sets. The proof involves showing that as long as the population is sufficiently large, it is much easier for mutations to introduce Altruists into a world of Egoists than for mutations to eradicate Altruists from a purely altruistic or mixed world.

One's initial impression might be that mutations should be inimical to Altruists, because a mutant Egoist will thrive and grow when introduced into a collection of Altruists while a lone Altruist will wither and die when introduced into a collection of Egoists. Notice, however, that a small clump of Altruists in the midst of Egoists will not only survive, but will grow. It takes only three adjacent Altruists in a world that is otherwise completely Egoists to ensure that the imitation dynamics lead to an absorbing set containing at least 60-percent Altruists.

It takes only a single mutation to introduce an Egoist into a world of Altruists. However, the resulting Egoist string can grow no longer than three. In light of this, consider an initial state that consists only of Altruists. A mutation creating an Egoist or a clump of Egoists will prompt imitation dynamics leading to an absorbing state with no more than three Egoists. To get additional Egoists, additional mutations are required. These mutations can lead to states where there are many small clumps of Egoists. As these clumps become more numerous, and hence closer together, additional mutations join together previously separated clumps of Egoists. But as long as there are still some strings of Altruists, these newly joined strings of Egoists will shrink, replacing two original strings with a new, shorter string (of length three or less) and ultimately *decreasing* the proportion of Egoists. In order to further increase the proportion of Egoists, mutations must simultaneously eliminate *all* strings of Altruists. But this requires a large number of mutations, if N is large, and hence is extraordinarily unlikely. If N exceeds 30, then it takes at least four mutations to eliminate all Altruists, which suffices for the result.

Altruists can thus invade a world of Egoists with only a local burst of mutation that creates a small string of Altruists, which will then subsequently grow to a large number of

Young (1993) in concentrating on arbitrarily small mutation rates. In biology, the concept of an evolutionarily stable strategy (John Maynard Smith [1982]) is built around the presumption that mutations are arbitrarily improbable compared to the forces of selection.

[17] If Q_{ij} is the probability that mutations change the state to j, given that the current state is i, then $\Gamma_{ij} = \Sigma_k P_{ik} Q_{kj}$.

Altruists. Mutations can create small pockets of egoism, but these pockets destroy one another if they are placed too close together, placing an upper bound on the number of Egoists that can appear. The only possibility for surpassing this bound lies in a "global" mutation combination that simultaneously attacks all strings of Altruists. Mutations thus lead much more readily to absorbing sets with Altruists than absorbing sets without them, and the limiting distribution concentrates all of its probability on the former. This reinforces our finding that absorbing sets containing Altruists are the limiting outcomes in the absence of mutations, as long as there is some initial probability of altruism in a large population. Our model thus differs from many mutation-counting analyses, in that our limiting distribution does *not* depend critically on highly improbable sequences of mutations and hence need not involve extraordinarily long waiting times.

If we consider large populations, mutations will ensure that there are more than 60-percent Altruists in the population, though less than 100 percent. The exact calculation of the limiting distribution is tedious, but we can establish some bounds. The calculation of these bounds is significantly simpler for the case of $\mu < 1$, though μ can be arbitrarily close to one.[18]

The argument proceeds by noting that if the number of Egoist strings is too small, then the Egoist strings will be far apart and most Altruist strings will be long. A mutation will then tend to strike in the midst of Altruists and create a new string of Egoists, increasing the number of Egoists. If there are many Egoists strings, then these strings will be relatively close together, separated by short Altruist strings. Mutations will then often strike sufficiently close to two Egoist strings as to give rise to imitation dynamics that merge the two Egoist strings, thereby reducing the number of Egoists. We thus expect a centralizing tendency in the number of Egoists. We have:

PROPOSITION 6: *Let $\mu < 1$. Then the limit of the limiting distribution, as the population size gets large, restricts probability to absorbing sets in which the proportion of Altruists is between 70 percent and 87 percent.*

IV. Larger Neighborhoods

We have assumed that agents interact only with their immediate neighbors. This section examines an extension of the model that allows us to make the following point: decreasing the cost of altruism can be bad for Altruists.

We consider the case where each Altruist contributes one unit of the public good to each of his *four* closest neighbors. Each agent observes his own payoff and that of his four closest neighbors, and then chooses the strategy from those played by this group with the highest average payoff. We say that neighborhoods are of "radius two" in this case.

As in the previous case, the cost of altruism plays a crucial role in shaping the results. We study two intervals for the parameter C, namely $(3/4, 5/6)$ and $(5/6, 1)$.[19] Changing the value of C within such an interval does not affect the outcome, while we shall see that the two intervals give different behavior.[20] We take $\mu = 1$ throughout.

The investigation of the model with neighborhoods of radius two and costs $C \in (5/6, 1)$ begins with a calculation of transition rules. The nontrivial conditions under which an Altruist will remain an Altruist are the following cases and their mirror images, where the Altruist in the center is the agent in question:

[18] The advantage of $\mu < 1$ is that all of the absorbing sets under the random imitation dynamics are then singletons. This makes it easier to do the necessary calculations (and reduces the number of calculations) of the probabilities that mutations transform a given absorbing set into the basin of attraction of another.

[19] When agents interacted only with their immediate neighbors, the only relevant cost consideration was whether C was larger or smaller than $1/2$.

[20] If $C > 5/4$, then altruism is so costly that only Egoists survive. The results for $1 < C < 5/4$ are qualitatively similar to those for $5/6 < C < 1$, with one quantitative difference noted below. Cost levels $1/2 < C < 3/4$ give results similar to those of $3/4 < C < 5/6$. Costs $C < 1/2$ give noticeably different and more complicated behavior that we do not investigate here. Cost levels that lie at the boundaries between these various intervals create complications arising out of cases in which the average payoffs to Altruists and Egoists are equal.

x stands for a strategy that could be either A or E:

(5)

Eaaa a xEEE

aaaa a EEEx

aaaa a aEEE.

The cases in which an Egoist will become an Altruist are the following, as well as their mirror images:

(6)

aaaE E EEEx

Eaaa E EEEE

aaaa E EEEx.

These allow us to prove:[21]

PROPOSITION 7: *Let neighborhoods be of radius two and let* $^5/_6 < C < 1$. *Then absorbing sets generically consist of (i) the state in which all agents are Egoists and the state in which all agents are Altruists, and (ii) sets containing states in which strings of Altruists of length five or more are separated by either strings of three E's or blinkers, where blinkers consist alternately of one E and five E's or consist alternately of two E's and six E's. With the exception of the state in which all agents are Egoists, the proportion of Altruists is at least* $^5/_9$.

The proof mimics that of Proposition (1) and is omitted. To obtain the minimal proportion of Altruists in the stable sets that contain Altruists, we note that we can pack blinkers that alternate between two and six E's next to each other with five A's between them, in the following way:

(7) $\cdots$ *aaaaaEEaaaaaEEEEEEaaaaa* $\cdots$

$\cdots$ *aaaEEEEEEaaaaaEEaaaaaaa* $\cdots$.

This guarantees a maximum of Egoists, and here we have a proportion of $10/18 = {}^5/_9$ Altruists.

For costs in the interval $(^3/_4, {}^5/_6)$, we have the following:

PROPOSITION 8: *Let neighborhoods be of radius 2 and let* $^3/_4 < C < {}^5/_6$. *Then generically, absorbing sets include (i) the state in which all agents are Egoists and all are Altruists, and (ii) sets containing states in which strings of three or more Altruists are separated by strings of exactly three Egoists, or by blinkers which alternate between one and five or between two and six Egoists.*[22] *Except for the state in which all agents are Egoists, the proportion of Altruists in an absorbing set is a least* $^1/_2$.

The proof again mimics that of Proposition (1). To obtain the lower bound on the number of Altruists, note that it is possible to arrange blinkers in the following way:

(8) $\cdots$ *aaaEaaaEEEEEaaaEaaaEEEEEaaaEaaaaa* $\cdots$

$\cdots$ *aEEEEEaaaEaaaEEEEEaaaEaaaEEEEEaaa* $\cdots$.

This has $^1/_2$ of the population as Altruists. There is no denser way to arrange blinkers.

In each case a result analogous to Proposition 2 holds, establishing that if agents' initial identities are independently determined and may be altruistic, then as the population grows, the probability of convergence to a state in which altruism survives approaches unity.

The lower bound on Altruists is lower for $C \in (^3/_4, {}^5/_6)$ than in the case of higher costs, being $^1/_2$ rather than $^5/_9$. In this sense, it can be disadvantageous for Altruists to have their altruism come too cheaply. The forces behind this result are revealed by comparing (7) and (8). Example (7) reflects the fact that when

[21] The "generically" in this statement allows us to avoid values of C within the interval $(^5/_6, 1)$ that create payoff ties between Altruists and Egoists. For $1 < C < {}^5/_4$ $(C \neq {}^7/_6)$, we have the same characterization of absorbing sets, except that blinkers must be separated by at least six Altruists, making the minimal percentage of Altruists 0.6.

[22] If two one/five blinkers are separated by a string of only three Altruists, then the blinkers must be out of phase, so that the state in which one of the blinkers has five Egoists is the state in which the other blinker has only one Egoist. A two/six blinker requires at least five Altruists on each side.

costs are relatively high, strings of Altruists must be at least five Altruists long in order for Altruists to receive payoffs high enough to survive. Example (8) reflects the fact that for lower costs, Altruists' payoffs are higher and shorter strings (of length three) of Altruists can survive. The imitation dynamics can then lead to outcomes in which islands of Egoists are separated by strings of only three rather than five Altruists, and hence a smaller proportion of Altruists.

V. Conclusion

We have shown that if players choose their strategies in games by imitating successful players, and if there is a local or neighborhood structure to both the interaction between agents and their learning, then altruistic behavior can survive.

Imitation and the local nature of the interactions are both important to this result. Best-response learning would immediately lead agents to adopt the dominant strategy of Egoist. Agents who choose their strategies by imitating others will imitate Altruists, but only if the latter happen to be earning relatively high payoffs. The role of the local interaction structure is to allow Altruists to huddle together in concentrated groups. The benefits of the public goods supplied by Altruists are then enjoyed primarily by Altruists, leading to higher payoffs than those of Egoists, who tend to be surrounded by other Egoists.

A group of Altruists is always a ripe target for invasion by a mutant Egoist, who will thrive on the public goods provided by the Altruists. For this reason Altruists can survive, but they generally cannot conquer. Instead, the Altruists will be riddled with pockets of Egoists. However, there are limits to the expansion of egoism. As more and more Egoists try to free ride on nearby Altruists, the payoffs of Egoists fall and imitators become Altruists. The result is the preservation of altruism in coexistence with egoism. We can hope for altruism, but not for a perfect world of altruism.

Attention is drawn to the importance of the mutations in a local interaction model by results from biological studies of group selection. Group selection models typically assume that agents are arranged in isolated groups created by either spatial separation or kinship relationships.[23] Payoffs are measured in terms of expected numbers of offspring, and the proportion of the population playing a relatively high-payoff strategy increases because agents playing that strategy have relatively large numbers of offspring. As in our model of learning by imitation, the important property of this reproduction-based dynamic process is that if the agents in a location tend to be predominately Altruists, then altruism can spread to nearby locations, as the concentrated Altruists earn high payoffs and hence produce many offspring that spill over into neighboring locations.

Kin selection theories are now widely accepted as explanations for some seemingly altruistic behavior.[24] However, group selection models that are not based on kinship relationships have been criticized [e.g., Williams (1966); Dawkins (1976)], initially because the mechanism that caused some groups to grow faster than others was not specified, and subsequently because the combination of small groups, rare mutations, and infrequent migration required to support altruism is thought to be implausible.[25]

[23] See, for example, Vero Copner Wynne-Edwards (1962, 1986), W. D. Hamilton (1964, 1972), George C. Williams (1966), Eshel (1972), David Sloan Wilson (1975, 1987), Dan Cohen and Eshel (1976), C. Matessi and S. D. Jayakar (1976), Maynard Smith (1976), and Eshel and L. L. Cavalli-Sforza (1982).

[24] For example, kin selection arguments have been used to explain the behavior of several species of tropical butterflies, some of which incur a cost to develop a bitter taste that discourages birds from preying on others [Lincoln Pierson Brower and Jane Van Zandt Brower (1964), Brower (1969), Woodruff W. Benson (1971), and Eshel (1972)]. Kinship relationships play an especially important role in explaining the behavior of the social insects.

[25] Robert Boyd and Peter J. Richerson (1985) suggest that group selection arguments may be applicable in explaining the evolution of altruistic behavior among humans. They examine a model in which people have a taste for conformity, so that altruism is a *strict* best response as long as sufficiently many other people cooperate. In biology, group selection arguments are often invoked to explain the inefficiency of weapons used in competition for mates, such as excessively branched or curved horns (Konrad Lorenz, 1963), and are used to explain self-imposed limits on reproductive ability when a population is threatened by overpopulation. Wynne-Edwards (1962) suggests examples of the latter phenomenon, and evidence

Our model differs from a typical biological model of group selection in that our agents are arranged in overlapping rather than isolated groups. This overlapping-neighborhoods structure opens the possibility that mutations introducing Egoists into our model can favor Altruists while being detrimental to Egoists, by disrupting and eliminating strings of Altruists and hence causing two groups of Egoists to join and then shrink to a single, small group. This contrasts with typical biological group selection models, where migration and especially mutation work relentlessly against altruism and one must struggle to find a plausible explanation for why the migration and mutation rates are sufficiently small to allow altruism to survive. In addition, mutations in our model can introduce Altruists in the midst of Egoists. In biological models with isolated groups, such a possibility is thought to involve mutation rates that are unrealistically high and group sizes that are unrealistically small. In the presence of overlapping groups, relatively small numbers of mutations allow Altruists to gain a local foothold from which they can spill over into nearby locations. We could thus reinterpret our analysis as a biological model with local interactions and dynamics based on reproduction and emigration, providing a new explanation for altruism that exploits the overlapping group structure.

Our model of agents occupying locations around a circle is very simple. What happens if they are placed in a plane, or in a higher-dimension structure? To gain some insight into these cases, recall that Altruists fare poorly when exposed to many Egoists, while Egoists fare well when exposed to many Altruists. Taking agents to be arranged along a circle ensures that any group of A's cannot have too many Altruists who are on the boundary and hence are exposed to Egoists, and ensures that any group of Egoists cannot have too many members exposed to Altruists. This in turn produces conditions under which Altruists are likely to thrive. Moving to the plane or to richer spaces raises the possibility that groups of Altruists will appear that are irregularly shaped and that expose virtually all of their members to Egoists. These Altruists may then not survive, and the persistence of altruism appears to be less certain.

The extensive simulations of Nowak and May (1992, 1993) and Nowak et al. (1994) suggest that in the absence of mutations, there are many initial conditions from which a significant proportion of Altruists persist. Once again, Egoists in their model do well in the midst of Altruists while Altruists do poorly in the midst of Egoists, and concentrated groups of Altruists can then expand. The dynamics are much more complicated than in our simple model, but altruistic behavior typically survives.

What if the game contains more than two strategies? Altruists can survive in our model because the Altruists near the end of a long string of Altruists earn higher payoffs than do the Egoists near the end of long strings of Egoists. When we extended the argument to general 2×2 games, the criterion for expansion turned out to be a mixture of efficiency and risk dominance. In games with more than two strategies, the criteria for whether strategy x can expand at the expense of y, i.e., for whether agents on the boundary between strings of x and y will switch to x, involve pairwise efficiency and risk-dominance considerations. If there is a strategy that is relatively efficient and does not fare too badly in pairwise risk-dominance comparisons with all other strategies, then the system can converge to states featuring primarily that strategy. However, cyclic behavior can also appear in which strategy x expands at the expense of y, y expands at the expense of z, and z expands at the expense of x.

Imitation may often be important, but adopting the strategy observed to earn the highest average payoff is a very simple decision process. What about imitation rules other than simply comparing average payoffs? For example, agents may base their choices not only on average payoffs but also on the number of agents they observe playing each strategy. This introduces elements familiar from the literature on strategies for the infinitely repeated prisoner's dilemma. Tit for tat, for example,

provided by Frank Fenner (1965) (Mixoma viruses, which infect rabbits), John J. Christian (1970) (rodents and small mammals), and Frank M. Stewart and Bruce R. Levin (1984) (bacterial viruses).

simply adopts the previous strategy it has observed. The implications once again depend upon the behavior of an agent located at the end of a string of similar agents. An Altruist at the boundary between sufficiently long strings of Altruists and Egoists observes equal numbers of Altruists and Egoists among her opponents, as does the adjacent Egoist. If observing the Altruists in this sample makes agents more likely to cooperate by being Altruists, then our results are reinforced. If observing the Egoists makes the agents more likely to retaliate by being Egoists, and if this is sufficient to overcome the average payoff advantages of Altruists, then our results will be reversed and altruism will vanish. The success of altruism then depends upon whether agents facing both Altruists and Egoists tend to see their glasses as half full or half empty.

A great deal of work remains to be done in extending the analysis to larger games as well as more complicated spatial structures and learning rules. It is clear, however, that dynamics driven by imitation can differ significantly from the familiar best-reply dynamics and that imitation coupled with local interactions opens the possibility for altruistic behavior to survive.

Appendix

PROOF OF PROPOSITION 1:

It is immediate that the states in which all agents are Altruists or all agents are Egoists are absorbing states, because imitation cannot introduce Egoists into a world in which there are only Altruists, or vice versa.

To find the remaining absorbing sets, consider what happens to a string of A's as the imitation dynamics proceed. From (2), any A string of length one immediately disappears. Similarly, if we have an A string of length two, the two A's in this string immediately become E's. In the process, however, the adjacent E's may switch to A's. What happens to these adjacent E's? There are four possibilities. The following transitions describe the fate of the E (the center agent in each case) that initially sits just to the left of the string of two A's. A similar analysis holds for the E on the right. An "x" holds the place of an agent whose type we do not have sufficient information to ascertain (see below).

EE	*E*	*aa*	*aE*	*E*	*aa*	*Ea*	*E*	*aa*	*aa*	*E*	*aa*
xE	*a*	*EE*	*xE*	*E*	*EE*	*EE*	*E*	*EE*	*xE*	*E*	*EE*
EE	*E*	*EE*	*xx*	*E*	*EE*	*EE*	*E*	*EE*	*xx*	*E*	*EE*

Moreover, the x's in the final line can be A's only if there existed a string of three or more A's to the left of our segment, to which these agents have now become attached. Hence, an A string of length two disappears after two periods without creating any new A strings.

What of A strings that are of length three or longer? From (1)–(2), the A's at the end of such string are the only potential candidates for becoming E's, and the only way that such a string can increase in length is for a single adjacent E at an end to change to A. Hence, such a string may undergo a change in length of $\{-2, -1, 0, 1, 2\}$. Because the string can increase in length only if it borders a segment of three E's [from (1)], the string cannot merge with any other A strings of length three or more. There are then only two possible fates for such a string. It can persist forever as a distinct string, perhaps varying in length, or its length can fall below three at some point, causing it to be eliminated within the next two periods without giving birth to new strings. We thus have that strings of A's can be destroyed but cannot be created.

Together, these results give: *There exists a time τ such that the number of A strings at time τ is less than or equal to the number of A strings of length three or more at time zero; the number of A strings in any subsequent period is equal to the number at time τ; and all A strings in subsequent periods are length three or longer.*

What can we say about E strings? First, notice that the number of A and E strings must be equal. Next, suppose that time τ has been reached, so that all A strings have length at least three. Then from (1), any E string the length of which is more than two declines in length by two, a string of length two retains its length, and a string of length one increases in length by two. Hence, we will eventually have Egoist strings of length two or blinkers, alternating between lengths one and three, but no longer strings, giving: *There exists a time τ after which the number of E strings is less than the number of E strings in the initial state and is constant, and E strings either remain at length two or alternate between lengths one and three.* This gives Proposition 1.1. It is now an easy calculation to check that since A strings occur in lengths at least three, and since E strings occur in either length two or alternations between length one and three, that the proportion of A's, if there are to be any A's at all, must be at least 0.6.

PROOF OF PROPOSITION 2:

It is immediate that the system must converge to a state containing Altruists, and hence a state containing at least 60-percent Altruists (by Proposition 1) if there exists a persistent string, and that the probability of a persistent string approaches unity as N gets large if initial identities are randomly, independently determined, with positive probability on Altruist. It then remains to verify the characterization of persistent strings. We examine the case of a string containing at least five adjacent A's. Showing that the remaining strings identified in (2.1) are persistent, and that any other string is eliminated by period three, involves straightforward variations on this argument. (Proposition 1 has already shown that every string of length two or less is eliminated within two periods.)

We show that a string of A's, the length of which is at least five, cannot disappear. In particular, we show that if there exists a string of five A's at time t, then either all five of these agents must also be Altruists at time $t + 1$ or they must all be Altruists at time $t + 2$. This holds regardless of the strategies played by other agents in the system.

Suppose we have a string of five or more A's bordered on each end by an E. Each of these two E's must have either an A or E on its other side. This gives us four possibilities to consider. First, suppose each E has an A on its other side. Then from (1)–(2), the system proceeds as follows:

$$\cdots aE \quad aaaaa \quad Ea\cdots$$

$$\cdots EE \quad EaaaE \quad EE\cdots$$

$$\cdots xE \quad aaaaa \quad Ex\cdots .$$

As usual, an x holds the place of an agent who may be either an Altruist or an Egoist. For convenience, the original string of five A's is separated by spaces. A similar result clearly holds if the original string contains more than five A's.

Alternatively, one of the E's on the end of the string of A's may have an E on its other side while the other may have an A on its other side. This gives us the following case and its mirror image:

$$\cdots aE \quad aaaaa \quad EE\cdots$$

$$\cdots EE \quad Eaaaa \quad xE\cdots$$

$$\cdots xE \quad aaaaa \quad xx\cdots.$$

Finally, the E's on both ends of the string of A's may be bordered by E's. Then we have:

$$\cdots EE \quad aaaaa \quad EE\cdots$$

$$\cdots Ex \quad aaaaa \quad xE\cdots .$$

In each case, the result is that any string of at least five Altruists persists.

PROOF OF PROPOSITION 3:

Let there be countably many agents, denoted by the integers. Consider the initial state, and suppose that, in this state, agent 0 is the rightmost agent of one of the following sequences of agents:

$$\begin{array}{ccc} aaaaaE & EEaaaaE & EEEaaaE \\ & aaaaEE & EEaaaEE \\ & & aaaEEE. \end{array}$$

Let such a string be called a "persistent-string-with-right-boundary." Then we know that contained within agents $\{-6, \cdots, -1\}$ is a persistent Altruist string.

We now let τ be an integer with the property that, in the initial state, the set of agents $\{0, \ldots, \tau\}$ contains a persistent string. Our task is to calculate an upper bound on the expected value of τ.[26] Because the identities of the agents in $\{0, 1, \cdots\}$ are determined independently, each with probability p of being an Altruist, we can describe the initial condition, and hence the expected value of τ, by defining a new Markov process as follows. Let there be 17 states, denoted by $\{1, 2, \ldots, 16, T\}$, where we think of the first 16 of these states as being associated with the following sequences:

1: *aE*	7: *aEaa*	13: *aEaaaE*
2: *EE*	8: *EEaa*	14: *aEaaaa*
3: *EEE*	9: *EEEaa*	15: *aEaaaEE*
4: *aEa*	10: *aEaaa*	16: *aEaaaaE*.
5: *EEa*	11: *EEaaa*	
6: *EEEa*	12: *EEaaaE*	

We define the state at step 0 as being one of states 1, 2, or 3, depending upon whether agents $-2, -1$, and 0 are characterized by xaE ($x \in \{a, E\}$), aEE, or EEE. We define the state at step $t > 0$ to be state T if any agent from the set $\{0, \ldots, t\}$ is the rightmost agent in one of the following sequences:

aaaaa	*EEaaaa*	*EEEaaa*
	aaaaEE	*EEaaaEE*
		aaaEEE.

If the state at step t is not T, then the state is given by $i \in \{1, \ldots, 16\}$ if agent t is the rightmost agent in the sequence corresponding to state i, and the same is true of no other $i \in \{1, \ldots, 16\}$ with a longer sequence. Intuitively, the system begins at step zero in one of states 1, 2, or 3, which are the sequences of Altruists and Egoists that must lie at the right end of a persistent-string-with-right-boundary. We then examine individuals 1, 2, 3, and so on, in each case using that individual's identity as either Egoist or Altruist to define a transition of the new Markov process. The latter enters state T (for "terminal") whenever an individual has been encountered who allows us to confirm the existence of a persistent string. States 1 through 16 are fragments of persistent strings.

The transition probabilities for this new Markov process are calculated on the basis of the assumption that agents in the initial state are independently chosen to be Altruists with probability p and Egoists with probability $1 - p$. It is then straightforward to calculate the expected number of steps to absorption in state T from each of states 1, 2, and 3 (see E. Seneta, 1981 Theorem 4.5). Numerical calculations produce a table matching that given in Proposition 3, where these numbers are the maximum of the expected number of steps from the three initial conditions given by states 1, 2, and 3.[27] These figures are upper bounds on the expected number of agents between the end of a persistent-string-with-right-boundary and the end of the next persistent string to the right. The boundary of the persistent-string-with-right-boundary may contain up to three Egoists, but a persistent string must contain at least three Altruists, so these numbers are also upper bounds on the number of agents between two persistent strings of Altruists.

PROOF OF PROPOSITION 4:

Let the types of players be denoted by 1 and 2. Let each player i have a set of players whom he potentially imitates, called his learning neighborhood, and a set of players with whom he interacts, called his interaction neighborhood. In particular, player i's imitation rule

[26] The assumption that N is infinite makes its appearance here. It allows us to calculate the expected value of τ while ignoring the possibility that the set of agents is exhausted before encountering another persistent string, with our search for such a string having taken us around the circle and back to our point of departure at agent 0.

[27] The details of these calculations are avaiable on request. Differences in the expected number of steps from these three initial conditions are small. The maximum such difference appears for large values of p, and is 2 for $p = 0.5$ but only 0.3 for $p = 0.1$.

adopt the strategy that receives the highest average payoff of the strategies represented in her learning neighborhood.[28] In our model of Altruists and Egoists, the interaction neighborhood of player i included his two nearest neighbors, while his learning neighborhood included these two nearest neighbors and himself. Let N_i^{L1} and N_i^{L2} be the sets of type-1 and type-2 players in agent i's learning neighborhood. Let N_i^{I1} and N_i^{I2} be the sets of type-1 and type-2 players in agent i's interaction neighborhood. Let n_i^{L1}, n_i^{L2}, n_i^{I1}, and n_i^{I2} be the numbers of players in the sets N_i^{L1}, N_i^{L2}, N_i^{I1}, and N_i^{I2}. Let P_i be player i's payoff, and let $n_i^{I1} + n_i^{I2} = n^I$ be the size of the interaction neighborhood. According to the imitation rule, player i will become type 1 if and only if:

$$\text{(A1)} \quad \frac{1}{n_i^{L1}} \sum_{j \in N_i^{L1}} P_j > \frac{1}{n_i^{L2}} \sum_{j \in N_i^{L2}} P_j.$$

When player j is of type s, then his payoff is given by

$$P_j = K_1 n_j^{I1} + K_2 n_j^{I2} - C_s$$

$$= (K_1 - K_2) n_j^{I1} + K_2 n^I - C_s,$$

and the imitation rule (A1) becomes:

$$\frac{K_1 - K_2}{n_i^{L1}} \sum_{j \in N_i^{L1}} n_j^{I1}$$

$$> (C_1 - C_2) + \frac{K_1 - K_2}{n_i^{L2}} \sum_{j \in N_i^{L2}} n_j^{I2}.$$

It is now obvious that when $K_1 - K_2 > 0$, the dynamics will be identical for all pairs of types (K_s, C_s) for which $(C_1 - C_2)/(K_1 - K_2)$ is the same.

PROOF OF PROPOSITION 5:

Let $\mathcal{E}$ denote the state in which all agents are Egoists. Let $\mathcal{A}$ be the state in which all agents are Altruists. Let $\mathcal{X}(n, m)$ denote the collection of absorbing sets with the property that in any state contained in such an absorbing set, at least some agents are Altruists and all agents are Altruists except n strings of Egoists of length two and m blinkers, where $n \geq 0$ and $m \geq 0$. We define $\mathcal{X}(n, m)$ only for values of (n, m) for which $\mathcal{X}(n, m)$ is nonempty. Then $\mathcal{A}$ is the unique element in $\mathcal{X}(0, 0)$ and every absorbing set other than $\mathcal{E}$ is contained in some $\mathcal{X}(n, m)$.

It suffices to show that $D(\mathcal{A}) < D(\mathcal{E})$, where D is defined in Lemma 3 of Samuelson (1994). For this, it suffices to show that:

- Three mutations suffice to transform $\mathcal{E}$ into a state in the basin of attraction, under the imitation process, of a state in $\mathcal{X}(n, m)$ for some (n, m).
- Given any absorbing set in $\mathcal{X}(n, m)$ with $(n, m) \neq (0, 0)$, there exists a state in the absorbing set which a single mutation can transform into a state in the basin of attraction, under the imitation process, of an absorbing set in $\mathcal{X}(n', m')$ with $n' + m' < n + m$ or with $n' < n$ and $m' \leq m + 1$.
- Given any state in any absorbing set in $\mathcal{X}(n, m)$ for any (n, m), it takes at least $N/10$ mutations to reach a state in the basin of attraction, under the imitation process, of $\mathcal{E}$.

A state is in the basin of attraction, under the imitation process, of an absorbing set, if the deterministic imitation process (without mutations) leads from the state to the absorbing set.

To establish the first condition, we need only note that if three mutations introduce three adjacent Altruists into state $\mathcal{E}$, then Proposition 2 ensures that we then have a state in the basin of attraction $\mathcal{X}(n, m)$ for some (n, m). To establish the second condition, consider an absorbing set S' in $\mathcal{X}(n, m)$. If $m > 0$, then we need only choose a state in S' which at least one blinker has only one Egoist. A mutation switching this Egoist to an Altruist then produces a state in absorbing set in $\mathcal{X}(n, m - 1)$. Hence, consider an absorbing set in $\mathcal{X}(n, 0)$. Now let a mutation switch an Egoist to an Altruist. The result is an isolated Egoist (that was adjacent to the Egoist affected by the mutation). The next iteration of the imitation

[28] Player i is excluded from his interaction neighborhood and included in his learning neighborhood. In the case of Altruists and Egoists, this simply reflects our choice to measure the costs of altruism as the net costs, after any benefits of one's own public good provision have been realized.

process will produce a string of three Egoists. If all Altruist strings are still of length at least three, then we have a blinker and a state in an absorbing set contained in $\mathcal{X}(n-1, 1)$. If instead at least one Altruist string is now of length only two, then (from the proof of Proposition 1) that string of Altruists will disappear, while no new string can appear, yielding a state in an absorbing set in $\mathcal{X}(n', m')$ with $n' + m' < n$.

Finally, we calculate a lower bound on the number of mutations required to convert a state in an absorbing set in $\mathcal{X}(n, m)$ into a state in the basin of attraction of $\mathcal{E}$. The mutations must eliminate all of the strings of Altruists in the original state. We first notice that in order to eliminate a string of A's of length k, we must have at least $[k/5]$—the integral value of $k/5$—mutations.[29] A lower bound on the number of mutations needed to eliminate all string of A's is then $N/10$, which arises in the case in which there are strings of A's of length nine (which are the longest that can still be eliminated by a single mutation) with blinkers at the end of the string, where the blinkers are in phase and there are nine Altruists in the string when each blinker consists of a single Egoist. For sufficiently large N, and in particular for N exceeding 30, this number exceeds three, giving the result.

PROOF OF PROPOSITION 6:

Fix the population size N. Let the Markov process induced by the imitation dynamics be $(S, \mathbf{P})$, where S is the state space and $\mathbf{P}$ is the transition matrix, and let the Markov process induced by the imitation-and-mutation dynamics by $(S, \boldsymbol{\Gamma})$, where $\boldsymbol{\Gamma}$ is the transition matrix. We say that an agent chosen to assess her strategy, under the random imitation dynamics, has "received the learn draw."

Step 1: This step shows that instead of examining $(S, \boldsymbol{\Gamma})$, we can work with a simpler Markov process $(K, \boldsymbol{\Delta})$. To construct the simpler process, we let $[N/5]$ denote the integral value of $N/5$ and let the state space $K = \{0, 1, \ldots, [N/5]\}$. We interpret a state $k \in K$ as identifying the number of Egoist strings in an absorbing state of $(S, \mathbf{P})$.[30] The transition matrix is $\boldsymbol{\Delta}$, where Δ_{ij} is the probability that a *single* mutation in $(S, \boldsymbol{\Gamma})$, followed by the imitation dynamics, leads from an absorbing set with i Egoist strings to an absorbing set with j Egoist strings. Notice that a mutation can create at most one new Egoist string or can destroy at most one string, and hence can cause the number of Egoist strings to change by at most one. The proportion of Altruists in the limiting distribution of $(K, \boldsymbol{\Delta})$ matches the proportion in the limiting distribution of $(S, \boldsymbol{\Gamma})$.

Step 2: We now examine $(K, \boldsymbol{\Delta})$. This is a birth-death process, since from state k, there is positive probability of moving only to states $k-1$, k, and $k+1$. The stationary distribution δ^* of a birth-death process must satisfy the detailed balance condition:

$$\text{(A2)} \qquad \frac{\delta^*(k)}{\delta^*(k+1)} = \frac{\Delta_{k+1,k}}{\Delta_{k,k+1}}.$$

To complete the proof, it suffices to show that there is $\varepsilon > 0$ such that for any N, if $2k/N \le 0.13$ (recall that each Egoist string contains two Egoists), then $\Delta_{k,k+1}/\Delta_{k+1,k} > 1 + \varepsilon$ and if $2k/N \ge 0.30$, then $\Delta_{k,k+1}/\Delta_{k+1,k} < 1 - \varepsilon$. In particular, this ensures (from A2) that the ratio $\delta^*(k)/\delta^*(k+1)$ is bounded below one when $2k/N \le 0.13$ and bounded above one when $2k/N \ge 0.30$. As N grows, the number of pairs $(k, k+1)$ with $2k/N \le 0.13$ and $2k/N \ge 0.30$, and hence the number of pairs for which these bounds on the stationary distribution hold, approaches infinity. This can occur only if the probability attached to

[29] This number is calculated by observing that if an Egoist is placed in the midst of a string of Altruists, the result is a blinker, with three Egoists in the next period. In order to eliminate a string of A's, enough Egoists must be inserted so that after a period has passed and each Egoist given rise to a string of three Egoists, with blinkers possibly also converting the A's at each end of the string into E's, all remaining strings of A of the original string must be at most of length two. This requires at least $[k/5]$ mutations.

[30] The details of this construction, as well as the calculations from Step 3, are available on request. Since each such string must contain at least two Egoists and must be separated from other Egoist strings by at least three Altruists, there can be at most $[N/5]$ such strings.

to states k such that $2k/N \in (0.13, 0.30)$ approaches unity.[31]

Step 3: This step verifies the required inequalities. Recall that absorbing states consist of strings of two Egoists separated by strings of three or more Altruists. We first calculate a lower bound on $\Delta_{k,k+1}$. A mutation creates a new string of Egoists with probability one if it converts to egoism an Altruist who is bordered by at least four Altruists on each side; with probability between zero and one if the Altruist is bordered by three Altruists on one side and at least four on the other; and otherwise with probability zero. In light of this, we can find a lower bound on the probability of increasing the number of Egoist strings by arranging agents so that there are eight Altruists between each Egoist string, leaving one longer string of leftover Altruists, and assuming that a mutation inserting an Egoist between three Altruists on one side and four on the other never creates a new string of Altruists. The probability of introducing a new Egoist string is then bounded below by the probability that a mutation strikes an agent more than four Altruists away from the end of the long string of Altruists, or

$$\text{(A3)} \qquad \underline{\Delta}_{k,k+1} = \frac{1}{N}(N - 10k).$$

A similar calculation shows that the probability of introducing a new Egoist string is maximized if strings of Egoists are separated by strings of only three Altruists, giving an upper bound of:[32]

$$\text{(A4)} \qquad \bar{\Delta}_{k,k+1} = \frac{1}{N}(N - 5k - 3).$$

We now turn to the probability of eliminating Egoist strings. An upper bound on the probability of eliminating such a string is:

$$\text{(A5)} \qquad \bar{\Delta}_{k,k-1} = \frac{1}{N}5k.$$

[31] See Ken Binmore and Samuelson (1997) for a similar argument.

[32] The "−3" reflects a three-Altruist buffer at both ends of any long string of A's in which a mutation cannot create a new string of Egoists.

A lower bound on the probability of eliminating Egoist strings is given by:

$$\text{(A6)} \qquad \underline{\Delta}_{k,k-1} = \frac{1}{N}\frac{8}{9}2k.$$

We use these calculations to obtain:

$$\frac{\underline{\Delta}_{k,k+1}}{\bar{\Delta}_{k+1,k}} = \frac{N - 10k}{5(k+1)} > 1 + \varepsilon$$

if $k/N \leq 0.065$ (and hence there are no more than 13-percent Egoists), N is sufficiently large, and $\varepsilon < 0.075$. Similarly,

$$\frac{\bar{\Delta}_{k,k+1}}{\underline{\Delta}_{k+1,k}} = \frac{N - 5k - 3}{2(k+1)}\frac{9}{8} < 1 - \varepsilon$$

if $k/N \geq 0.15$ (and hence there are at least 30-percent Egoists), N is sufficiently large, and $\varepsilon < 0.06$. This gives the result.

REFERENCES

Benson, Woodruff W. "Evidence for the Evolution of Unpalatability by Kin Selection in the Heliconiinae (Lepidoptera)." *American Naturalist*, May–June 1971, *105*(943), pp. 213–26.

Bergstrom, Theodore C. and Stark, Oded. "How Altruism Can Prevail in an Evolutionary Environment." *American Economic Review*, May 1993 (*Papers and Proceedings*), *83*(2), pp. 149–55.

Binmore, Ken and Samuelson, Larry. "Muddling Through: Noisy Equilibrium Selection." *Journal of Economic Theory*, June 1997, *74*(2), pp. 235–65.

Blume, Lawrence E. "The Statistical Mechanics of Strategic Interaction." *Games and Economic Behavior*, July 1993, *5*(3), pp. 387–424.

Boyd, Robert and Richerson, Peter J. *Culture and the evolutionary process*. Chicago: University of Chicago Press, 1985.

Brower, Lincoln Pierson. "Ecological Chemistry." *Scientific American*, February 1969, *220*(2), pp. 22–29.

Brower, Lincoln Pierson and Brower, Jane Van Zandt. "Birds, Butterflies and Plant Poisons: A Study in Ecological Chemistry."

Zoologica, November 2, 1964, *49*(3), pp. 137–59.

Christian, John J. "Social Subordination, Density and Mammalian Evolution." *Science*, April 3, 1970, *168*(3927), pp. 84–90.

Cohen, Dan and Eshel, Ilan. "On the Founder Effect and the Evolution of Altruistic Traits." *Theoretical Population Biology*, December 1976, *10*(3), pp. 276–302.

Dawkins, Richard. *The selfish gene*. Oxford: Oxford University Press, 1976.

Ellison, Glenn. "Learning, Local Interaction, and Coordination." *Econometrica*, September 1993, *61*(5), pp. 1047–72.

Eshel, Ilan. "On the Neighbor Effect and the Evolution of Altruistic Traits." *Theoretical Population Biology*, September 1972, *3*(3), pp. 258–77.

Eshel, Ilan and Cavalli-Sforza, L. L. "Assortment of Encounters and Evolution of Cooperativeness." *Proceedings of the National Academy of Sciences*, February 1982, *79*(4), pp. 1331–35.

Fenner, Frank. "Myxoma Virus and Cryctologus Cuniculus: Two Colonizing Species," in H. G. Baker and G. Ledyard Stebbins, eds., *The genetics of colonizing species*. New York: Academic Press, 1965, pp. 485–99.

Freidlin, M. I. and Wentzell, A. D. *Random perturbations of dynamical systems*. New York: Springer-Verlag, 1984.

Fudenberg, Drew and Maskin, Eric. "The Folk Theorem in Repeated Games with Discounting and Incomplete Information." *Econometrica*, May 1986, *54*(3), pp. 533–54.

Gilboa, Itzhak and Schmeidler, David. "Case-Based Decision Theory." *Quarterly Journal of Economics*, August 1995, *110*(3), pp. 605–40.

———. "Case-Based Optimization." *Games and Economic Behavior*, July 1996, *15*(1), pp. 1–26.

Hamilton, W. D. "The Genetic Evolution of Social Behavior." *Journal of Theoretical Biology*, January 1964, *7*(1), pp. 1–52.

———. "Altruism and Related Phenomena." *Annual Review of Ecological Systems*, 1972, *3*, pp. 193–232.

Huberman, Bernardo A. and Glance, Natalie S. "Evolutionary Games and Computer Simulations." *Proceedings of the National Academy of Sciences,* August 15, 1993, *90*(16), pp. 7716–18.

Kandori, Michihiro; Mailath, George J. and Rob, Rafael. "Learning, Mutation, and Long Run Equilibria in Games." *Econometrica*, January 1993, *61*(1), pp. 29–56.

Kemeny, John G. and Snell, J. Laurie. *Finite Markov chains*. Princeton, NJ: D. Van Nostrand, 1960.

Kreps, David M.; Milgrom, Paul; Roberts, John and Wilson, Robert J. "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma." *Journal of Economic Theory*, August 1982, *27*(2), pp. 245–52.

Lorenz, Konrad. *On aggression*. Harcourt, Brace & World, 1963.

Matessi, C. and Jayakar, S. D. "Conditions for Altruism Under Darwinian Selection." *Theoretical Population Biology*, June 1976, *9*(3), pp. 360–87.

Maynard Smith, John. "Group Selection." *Quarterly Review of Biology*, June 1976, *51*(2), pp. 277–83.

———. *Evolution and the theory of games*. Cambridge: Cambridge University Press, 1982.

Nowak, Martin A.; Bonhoeffer, Sebastian and May, Robert M. "More Spatial Games." *International Journal of Bifurcation and Chaos*, February 1994, *4*(1), pp. 33–56.

Nowak, Martin A. and May, Robert M. "Evolutionary Games and Spatial Chaos." *Nature*, October 29, 1992, *359*, pp. 826–29.

———. "The Spatial Dilemmas of Evolution." *International Journal of Bifurcation and Chaos*, February 1993, *3*(1), pp. 35–78.

Samuelson, Larry. "Stochastic Stability in Games With Alternative Best Replies." *Journal of Economic Theory*, October 1994, *64*(1), pp. 35–65.

Seneta, E. *Non-negative matrices and Markov chains*. New York: Springer-Verlag, 1981.

Stewart, Frank M. and Levin, Bruce R. "The Population Biology of Bacterial Viruses: Why Be Temperate?" *Theoretical Population Biology*, August 1984, *26*(1), pp. 93–117.

Trivers, Robert L. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology*, March 1971, *46*(1), pp. 35–57.

Williams, George C. *Adaptation and natural selection*. Princeton, NJ: Princeton University Press, 1966.

Wilson, David Sloan. "A Theory of Group Selection." *Proceedings of the National Academy of Sciences*, January 1975, *72*(1), pp. 143–46.

______. "Altruism in Mendelian Populations Derived from Sibling Groups: The Haystack Model Revised." *Evolution*, September 1987, *41*(5), pp. 1059–70.

Wynne-Edwards, Vero Copner. *Animal dispersion in relation to social behavior*. Edinburgh: Oliver and Boyd, 1962.

______. *Evolution through group selection*. Oxford: Blackwell Scientific Publications, 1986.

Young, H. Peyton. "The Evolution of Conventions." *Econometrica*, January 1993, *61*(1), pp. 57–84.

When Does It Take a Nixon to Go to China?

By Alex Cukierman and Mariano Tommasi*

Substantial policy changes (like market-oriented reforms by populist parties and steps towards peace by "hawks") are sometimes implemented by "unlikely" parties. To account for such episodes this paper develops a framework in which incumbent politicians have better information about the state of the world than voters. The incumbent is unable to credibly transmit ***all*** *this information since voters are also imperfectly informed about his ideology. This paper identifies conditions under which an incumbent party's electoral prospects increase the more atypical the policy it proposes. Popular support for a policy, or its "credibility," depends on the policy maker-policy pair.* (*JEL* D7, D8, E6, C72)

The history of public policy contains several episodes in which structural reforms or important economic or foreign policy shifts were implemented by parties or policy makers whose traditional position was to oppose such policies. Argentina under (Peronist) Menem, Peru under Fujimori, and Bolivia under (populist) Paz Estenssoreo underwent profound market-oriented economic reforms.[1] France privatized some of its public sector and shifted the emphasis of economic policy to price stability during the 1980's, under socialist President Mitterand. In the late 1970's, after years of vehement opposition to trading land for peace, hawkish Israeli Prime Minister Begin yielded the entire Sinai peninsula in return for peace with Egypt. His partner to this historical deal was President Sadat of Egypt who is considered to be the first Arab leader to mount a relatively effective military campaign against Israel. Having established a strong and persistent anti-Communist record during the 1950's and 1960's, President Nixon then opened the door to the international legitimization of the People's Republic of China in the early 1970's.

These episodes should not be interpreted to imply that large shifts in policy can be implemented *only* by political parties having an historical bias against such policies. Privatization and other reforms under Thatcher are an obvious counterexample. But the examples in the preceding paragraph raise an intriguing and important question about the circumstances under which policies are implemented by "unlikely" political parties rather than by parties the ideologies of which favor such policies. Our objective in this paper is to identify conditions under which shifts in policy are more likely to be implemented by improbable characters. To this end, we first develop a political economy framework in which such a phenomenon *can* occur and use it to pin down a set of conditions which makes it more likely that such policies will be implemented by the "wrong" parties.

We then develop an explanation for "policy reversals" within the framework of asym-

* Cukierman: Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv 69978, Israel, and Center for Economic Research, Tilburg University, Tilburg 5000 LE, The Netherlands; Tommasi: Universidad de San Andrés (1644) Victoria, Buenos Aires, Argentina. The Amnon Ben Nathan Chair in Economics at Tel Aviv University, the Harvard/MIT RTG in Positive Political Economy and CIBER, and the Academic Senate at UCLA provided financial support. We benefited from useful suggestions of three anonymous referees. We are indebted to Olivier Blanchard, Greg Hess, Eric Rasmusen, Tom Piketty, and seminar participants at Boston College, Brown University, the University of Chicago, Dartmouth College, the European Center for Advanced Research in Economics at the Universite Libre de Bruxelles, the University of Geneva, Groningen University, the Innocenzo Gasparini Institute for Economic Research at Bocconi University, the University of Limburgh, MIT, the Universidad de San Andrés, the University of Strathclyde, Tel Aviv University, the University of Tilburg, and UCLA for helpful comments, and to Carola Schenone for research assistance.

[1] Dani Rodrik (1993) points out that it is ironic that populist and interventionist parties have implemented radical trade liberalizations, fiscal adjustments, and market-oriented reforms.

ic information about the mapping of policy instruments into policy outcomes.[2] Incumbent politicians normally have better information than the general public about the likely outcomes of alternative policies. Governments deal with public policy issues on a daily basis, they have access to the advice of specialists, and in some cases, they possess classified information. Furthermore, since information collection is costly, a large part of the voting public has neither the incentive nor the ability to become fully informed about all the aspects of complex public policy issues.[3] This idea is captured here by assuming that a key stochastic parameter of the mapping of policies into outcomes is observed by policy makers in office but not by the voting public.

People's welfare depends on outcomes, and policy choices affect outcomes. However, outcomes are also influenced by external circumstances about which policy makers normally have better information than does the general public. Thus, depending on external circumstances, a right-wing policy may or may not be desirable from the point of view of a majority of the population. Suppose that it is, and that the incumbent party is fully informed about this. In order to implement a policy, the party in office has to elicit support. To that end, it has to transmit to the public its private information about the relative desirability of, in this case, right-wing policies. When the incumbent is a recognized left-winger, his ability to do so and to implement the required policies is greater than the ability of a right-wing incumbent. The reason for this apparent incongruity is that the public has less reason to suspect that the right-wing policy is proposed solely because of the natural ideological tendencies of the party in office, i.e., it may be perceived as an objectively motivated policy.

Another important element of our framework is the inability of voters to fully distinguish policy shifts that are due to changes in (information about) the mapping of policy into outcomes from those that are due to intraparty politics.[4] This inability prevents the public from making precise inferences about government's private information concerning this mapping (the state of the world). The reason is that the policy proposals of incumbent governments are affected by shocks to this mapping as well as by changes in the incumbent party's preferred policy position. The second type of change makes the proposed policy a noisy indicator for the state of the world.

Some conditions make a policy reversal more likely. The policy switch that—in view of external circumstances—is desirable should be considerable *and* relatively rare. These conditions appear to have been satisfied in the episodes mentioned before. Major economic reforms, the trading of land for peace, and the opening of a pathway towards China are policy decisions of a scope that occurs infrequently. Furthermore, policy reversals are more likely when the voting public has substantial uncertainty about the governments' exact preferences. In addition, the final outcomes of the policies under consideration occur far in the future—that is why the voting public has to use the policy proposals of incumbents as signals for the likely future outcomes of the proposed policies. This feature is formalized by assuming that voting takes place *before* the realization of final outcomes.[5]

The basic ideas of the paper are presented in Section I with a representative-democracy model in which, after receiving some information

[2] The notion that this mapping is stochastic is not new. Recent references are Thomas Gilligan and Keith Krehbiel ([illegible]), Alberto Alesina and Cukierman (1990), Kenneth Rogoff (1990), Joseph E. Harrington, Jr. (1993), John Roemer (1994), and Christian Schultz (1996). David Austen-Smith (1993) emphasizes that legislation (or policy more generally) is a means to an end rather than a final outcome. Several of these authors also postulate, as we do, that incumbent parties have better information about some aspects of the mapping of policies into outcomes than does the general public.

[3] This naturally leads to specialization in knowledge (different people know more about some things than others) and to rational ignorance. This idea goes back to Anthony Downs (1957); for recent treatments see Arthur Lupia ([illegible]), Gilligan (1993), and John Matsusaka (1995).

[4] Michael Laver and Norman Schofield (1990) have stressed the effects of intraparty politics on policy choices.

[5] This timing is a crucial difference between our model and that in Harrington (1993). Harrington uses an otherwise similar informational structure to derive implications from voter uncertainty to policy manipulation for reelection purposes.

about the mapping of policies into outcomes, the incumbent party commits to a policy platform.[6] This is followed by elections. If the incumbent is reelected, he implements the proposed policy; if another party is elected, it picks the policy that is nearest to its own preferences given the realization of the stochastic component of the policy-to-outcome mapping.[7]

We show in Section II that *moderate* right-wing policies are more likely to be implemented by right-wing parties (and similarly for the left), but *extreme* right-wing policies are more likely to be implemented by left-wing parties (and vice versa). Section III contains comparative statics. In Section IV, we discuss some implications for credibility. Section V concludes, pointing to the next steps in this research agenda.

I. The Model

The economy consists of a large number of individual voters with different preferences over a single policy issue. The utility of a type j voter is given by:

$$(1)\qquad -|x^e - (c_j + \gamma^e)|$$

where x is policy, c_j is a constant, and γ is a normally distributed stochastic variable with zero mean and variance σ^2_γ. The superscript e attached to x and γ denotes expected values of these variables conditioned on the information available to the voter. $(c_j + \gamma^e)$ is the (perceived) ideal policy of a type j voter. It depends on the type-specific "taste" parameter c_j, as well as on the voter's perception of the realization of an exogenous state-of-nature parameter, γ, that induces a unidirectional shift in the preferred policies of all voters.

γ is meant to capture the effect of external circumstances on the ideal *policies* of voters. Voters have well-defined and stable preferences over outcomes, but the mapping of policies into outcomes, and hence the individual's utility function over policies, has a stochastic element.

People generally have different preferences or opinions about the desirability of alternative policies. But when they learn that there has been a shift in exogenous circumstances, all voters shift their preferred policies in the same direction. To illustrate, consider the "land versus peace" issue in the ongoing negotiations between Israel and Syria. Different individuals in Israel have different opinions about how much territory to give up for a peace of a given quality; some would give up very little and others a lot. In the language of economics, they possess different marginal rates of substitution. But the policy preferred by each individual (how much land to give up) also depends on his or her perception of the "deal" that Israel can work out with Syria. Again in the language of economics, preferred policies also depend on the exogenous marginal rate of transformation between land and peace. When they believe that it is possible to obtain a higher-quality peace for a given amount of territory, *all Israelis* advocate more dovish policies, although hawks are still willing to give up less territory than doves. The heterogeneity of preferences is captured by c_j and the common effect of perceived external circumstances by γ^e.

In a representative democracy, voters do not choose policy directly. Instead, they choose elected officials who decide what policy to follow. We model this institutional setup by postulating two parties, the right R and the left L, that compete for office. Each party cares about the issues as well as about being in office per se (this Downsian component is called ego rents by Rogoff and Anne Sibert, 1988). We use h to denote this "value of office"

[6] We are assuming that the incumbent takes an action that commits him to a future policy (like making a statement or sending a bill to Congress) thereby revealing part or all of his private information to the public. Although the action taken by the policy maker might include a verbal statement, it is not "cheap talk" in the sense of Vincent Crawford and Joel Sobel (1982). As stressed by Cukierman and Nissan Liviatan (1991), announcements of future policies by incumbent politicians are not necessarily costless from their point of view. Thus, although our analytical structure bears some resemblance to models of information transmission in debates—such as those in Austin-Smith (1990, 1992)—it is based on costly rather than costless signalling.

[7] In Cukierman and Tommasi (1998), we explore the same ideas in the context of a referendum game (direct democracy). The broad conditions for the type of reversal discussed in this paper apply to both institutional structures.

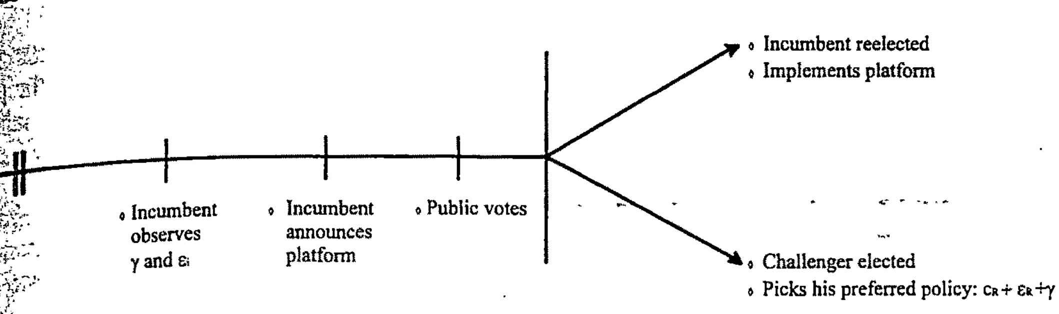


FIGURE 1. SEQUENCE OF EVENTS

and c_i to denote the (deterministic part of the) bliss point of party $i = L, R$, with $c_L < c_R$.

An idea well established in political science—see Otto Kirchheimer (1966) and Laver and Schofield (1990)—is that different parties cater to the interests of different but contiguous groups of constituencies. The ideal policy of the left-wing party reflects a compromise between the different leftist groups, and similarly for the right-wing party. The relative ability of each such group to affect the party's policy position is usually in a state of flux and not fully known by the general public. We model this effect by postulating that the preferred policies of parties are subject to intraparty shocks labelled ε_i, $i = L, R$. The realizations of these shocks, which are independently and normally distributed with mean zero and variance σ_ε^2, are known to the parties but not to the public. In addition, as is the case with voters, the parties' ideal points also depend on γ. Those factors are incorporated by postulating that the bliss point of party i is given by $(c_i + \varepsilon_i + \gamma)$, and its objective function by

$$h - |x_i - (c_i + \varepsilon_i + \gamma)|.$$

The deterministic component, c_i, is common knowledge, but the stochastic component, ε_i, is not known by the public. Similarly, the realization of γ is observed by the incumbent party, but not by the general public. This is a stylized way of expressing the presumption that government has a more precise notion than does the general public about the effect of external circumstances on the way in which policy instruments map into outcomes. Since the incumbent party has its own policy preferences, it is not able to transmit fully this information to voters. But the voting public normally learns something about the realization of γ from the incumbent's policy proposal.

A. *Timing, Information, and Elections*

It is convenient to divide the sequence of events, illustrated in Figure 1, into two periods. At the beginning of the first period, the incumbent observes γ and his own ε. He then makes a proposal x. After having observed the incumbent's proposal, the general public votes for or against the incumbent. In the second and last period, public policy is carried out. If the incumbent is reelected, he carries out the policy proposed in the first period (the proposal is a binding commitment). If the challenger gets elected, he picks the policy that maximizes the *ex post* value of his objectives (knowing γ and his own ε).

This way of modeling the interaction between the voting public and government is designed to capture the fact that a policy has to satisfy two conditions in order to be implemented. First, the government has to propose the policy; second, the proposal should draw sufficient public support. Obviously, the model can be interpreted literally as stated in the previous paragraph. But it is also possible to interpret the model more broadly, as capturing the fact that a policy cannot be implemented if it does not gather sufficient public support or if it raises too much opposition. Under such a broader interpretation, a "vote against the incumbent" would refer to a state in which the proposed policy is abandoned after some trial period for lack of sufficient

public support (even in the absence of formal elections), while a "vote for the incumbent" permits the continuation of the policy.[8]

Note that there is an asymmetry between the incumbent and the challenger. While the first commits to a policy prior to elections, the second—if elected—gets to choose his preferred policy. This asymmetry reflects the presumption that reputational and other considerations make it more difficult to adjust policy for the incumbent than for the challenger.[9]

At election time, each individual votes for the party that he perceives is going to implement a policy nearest to his own preferred policy. But this depends, in turn, on the individual's perception of the exogenous shock γ. Knowing that the incumbent's proposed policy partially reflects that party's private information about γ, each voter utilizes governmental policy as a signal for γ (in a way that will be made explicit below).

Let $g(c_j, \omega)$ be the fraction, or density, of voters with preference parameter c_j, where ω is a stochastic shift parameter. Given the realization of ω, the distribution of c_j is nonstochastic, so that electoral uncertainty is captured by the distribution of ω. Each realization of ω induces a different nonstochastic distribution of c_j, each with its own median c_m. The median c_m is, hence, a function of ω; we assume that it is a linear function, and that ω has a uniform distribution. This implies that c_m also possesses a uniform distribution. Let $\bar{c}$ and $\underline{c}$ be, respectively, the upper and lower bounds of this distribution.

Since voters' preferences are single-peaked, the outcome of the elections is determined by the preferred policy of the median voter. The party whose expected policy position is nearest to the ideal policy $(c_m + \gamma^e)$ of the median voter on election day wins the elections. Uncertainty about the distribution of voters' ideal points induces probabilistic voting—the probability of reelection of an incumbent party is, in general, strictly between zero and one (see Peter J. Coughlin, 1992).[10]

B. Equilibrium

Suppose, for concreteness and without loss of generality, that the incumbent is party L. Since policy must be chosen prior to elections and since the election outcome is uncertain, the incumbent party takes into consideration the effect of the current policy choice on its probability of reelection. More precisely, it chooses policy x_L so as to maximize:

$$(2)\quad P^L(x_L)[h - |x_L - (c_L + \varepsilon_L + \gamma)|] + [1 - P^L(x_L)] \times [-|E(x_R|\gamma) - (c_L + \gamma)|].$$

$P^L(\cdot)$ is the probability that L will be reelected and depends on policy choice x_L. The second term in the incumbent party's objective is the utility it expects to obtain if it loses the elections and policy is chosen by the challenger.[11] As the incumbent knows γ, the policy it expects from R is $E(x_R|\gamma)$.

The functional form of $P^L(\cdot)$ depends on the way the voting public forms its perception of γ. However, the formation of this perception depends on the (optimal) policy rule of the incumbent which depends, in turn, on $P^L(\cdot)$. In equilibrium, the policy rule that the public postulates in order to form perceptions of γ has to be consistent with the actual policy rule followed by the incumbent, and the public's

[8] Policy is known to react to opinion polls even between elections. See Benjamin Page and Robert Shapiro (1983) for U.S. evidence. These considerations are developed in more detail in Cukierman and Tommasi (1998), with a focus on developing countries.

[9] Daniel Ingberman (1984) and M. Daniel Bernhardt and Ingberman (1985) analyze the notion that the platforms of incumbents are more reliable indicators of their future policies than are the platforms of challengers. Schultz (1996) discusses the case in which *both* parties commit to their respective policy platform.

[10] It is commonly accepted that probabilistic voting is more realistic than deterministic voting. For example, Randall Calvert (1986 p. 54) points out that "this view is in harmony with the importance attached by traditional political scientists to the role of imperfect information in politics."

[11] We assume that the intraparty shock ε_L is "persistent." That is, it persists into the postelection period only if the incumbent party is reelected. Besides simplifying the algebra considerably, this assumption has some descriptive realism since success in the elections normally increases the durability of a given deviation from the party's bliss point, c_L, while failure normally reduces it.

expectation formation process assumed by the incumbent has to be identical to the actual process of expectation formation. A full definition of equilibrium follows.

Definition 1: An equilibrium is a pair of policy functions, $x_L(\gamma, \varepsilon_L)$ and $x_R(\gamma, \varepsilon_R)$, together with voters' beliefs, $\gamma^e(x_L)$, such that:

- The incumbent party chooses policy (prior to elections) so as to maximize (2).
- If elected, the challenging party chooses policy after elections so as to minimize $|x_R - (c_R + \varepsilon_R + \gamma)|$.
- Given his perception of γ (and of the policy of the challenging party), the median voter votes for the party whose expected policy is nearest to his ideal point.
- Voters' perceptions about γ (and about the policy of the challenging party, if elected) are formed rationally, using all the available information.

If the right-wing challenger is elected, he selects the policy that maximizes his *ex post* objectives, namely, he implements policy:

$$(3) \qquad x_R = c_R + \varepsilon_R + \gamma.$$

The policy expected by voters from R, prior to elections is, therefore:

$$(4) \qquad x_R^e \equiv E(x_R | x_L) = c_R + \gamma^e.$$

Note that the policy expected from R depends on the policy proposed by the left-wing incumbent prior to elections. The reason is that the choice of policy by both R and L depends on γ, and that voters receive information about γ from x_L.

The choice of policy by the left-wing incumbent is more complicated, since he has to take into consideration the effect of his current policy on voters' expectations and, through them, on the probability of reelection. Using (3) and the fact that $E(\varepsilon_R) = 0$ and that $E(x_R|\ldots) = c_L + \gamma$, the objective function (2) can be written as:

$$P^L(x_L)[h - |x_L - (c_L + \varepsilon_L + \gamma)|] - [1 - P^L(x_L)](c_R - c_L).$$

The equilibrium solution for x_L is obtained by the method of undetermined coefficients. We first postulate that the equilibrium choice of x_L is the following linear function of γ and ε_L:

$$(6) \qquad x_L = B_L + b_{L\gamma}\gamma + b_{L\varepsilon}\varepsilon_L$$

where B_L, $b_{L\gamma}$, and $b_{L\varepsilon}$ are coefficients to be determined. Observation of x_L by voters does not enable them to disentangle the effect of γ from the effect of ε_L; it is easy to verify that this implies that $b_{L\gamma} = b_{L\varepsilon} = b_L$. Therefore, (6) simplifies to

$$(7) \qquad x_L = B_L + b_L(\gamma + \varepsilon_L).$$

Voters know the decision rule in (7), observe x_L prior to elections, and use it to improve their forecast of γ. Since B_L is a known combination of parameters, it is easy to show that the expected value of γ, conditional on x_L, is given by (Morris De Groot, 1970 p. 169):

$$(8) \qquad \gamma^e \equiv E(\gamma | x_L) = \theta \frac{(x_L - B_L)}{b_L}$$

where

$$(9) \qquad \theta \equiv \frac{\sigma_\gamma^2}{\sigma_\gamma^2 + \sigma_\varepsilon^2}.$$

Since the policy of any incumbent becomes more right-wing the larger the γ (it will be shown that $b_L > 0$)—and this is known by the public—a more right-wing policy (a larger x_L) is *partially* interpreted by the public as being due to a larger γ. But this signal is noisy because the incumbent's policy choice is also influenced by the intraparty shock. For this reason, the impact of x_L on γ^e is weighted by θ.

ASSUMPTION 1: *We assume that voters believe that the policy proposed by the incumbent L is always to the left of that of the challenger R, or:* $x_L^e = x_L < x_R^e$. *The assumption states that party R is always perceived to be the right-wing party. Conditions on the model's parameters for the fulfillment of this assumption are presented in the*

Appendix (see also Alesina and Cukierman, 1990).

Assumption 1, together with the decision rule of the median voter, implies that there exists a critical value of c_m, denoted c_m^c, such that if $c_m \leq c_m^c$, the incumbent L wins the elections, and if $c_m > c_m^c$, the challenger R wins the elections. The value of c_m^c is obtained from:

$$|x_L - \gamma^e - c_m^c| = |x_R^e - \gamma^e - c_m^c|,$$

which, due to Assumption 1, is equivalent to:

$$(10) \quad (\gamma^e + c_m^c) - x_L = x_R^e - (\gamma^e + c_m^c).$$

Rearranging (10) and using (4), we obtain:

$$(11) \qquad c_m^c = \tfrac{1}{2}(c_R + x_L - \gamma^e).$$

The probability $P^L(x_L)$, that the left-wing incumbent is reelected, is equal to the probability that the ideal point, c_m, of the stochastic median voter falls to the left of c_m^c. Since c_m is uniformly distributed, this probability is given by[12]:

$$(12) \qquad P^L(x_L) = \frac{c_m^c - \underline{c}}{\bar{c} - \underline{c}}.$$

Using (8) in (11) and the resulting expression in (12), we obtain:

$$(13) \quad P^L(x_L) = \frac{1}{2(\bar{c} - \underline{c})}\left[c_R - 2\underline{c} + \theta\frac{B_L}{b_L} + dx_L\right],$$

where $d \equiv 1 - \theta/b_L$.

Substituting (13) into L's objective function (5), we obtain the following first- and second-order conditions for an internal maximum:

Case 1: If $x_L > c_L + \varepsilon_L + \gamma$, the first- and second-order conditions are given respectively by:

$$(14) \qquad d(h - x_L + \varepsilon_L + \gamma + c_R) - 2(\bar{c} - \underline{c})P^L(x_L) = 0,$$

$$(15) \qquad -d < 0.$$

Case 2: If $x_L < c_L + \varepsilon_L + \gamma$, the first- and second-order conditions are given respectively by:

$$(16) \quad d[h + x_L - (2c_L + \varepsilon_L + \gamma) + c_R] + 2(\bar{c} - \underline{c})P^L(x_L) = 0,$$

$$(17) \qquad d < 0.$$

Rearranging the first-order conditions, we obtain:

Case 1: $(x_L > c_L + \varepsilon_L + \gamma)$.

$$(18) \quad x_L = \frac{1}{2d}\left[d(c_R + h) - c_R - \theta\frac{B_L}{b_L} + 2\underline{c}\right] + \frac{1}{2}(\gamma + \varepsilon_L).$$

Case 2: $(x_L < c_L + \varepsilon_L + \gamma)$.

$$(19) \quad x_L = \frac{1}{2d}\left[2\underline{c} - c_R - \theta\frac{B_L}{b_L} + d(2c_L - c_R - h)\right] + \frac{1}{2}(\gamma + \varepsilon_L).$$

Equating coefficients across equations (18), (19), and (7) we obtain:

$$(20) \qquad b_L = \tfrac{1}{2}$$

[12] Obviously, the probability function is bound to belong to [0, 1]. In Part 2 of the Appendix, we analyze what happens when $c_m^c < \underline{c}$ and when $c_m^c > \bar{c}$.

and

$$(21)\quad B_L = \begin{cases} \dfrac{1}{2}\left[(2\underline{c} - c_R)\left(1 + \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right) + (h + c_R)\left(1 - \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)\right] & \text{for } x_L > c_L + \varepsilon_L + \gamma \text{ (Case 1).} \\[2ex] \dfrac{1}{2}\left[(2\underline{c} - c_R)\left(1 + \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right) + (2c_L - c_R - h)\left(1 - \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)\right] & \text{for } x_L < c_L + \varepsilon_L + \gamma \text{ (Case 2).} \end{cases}$$

Equation (20) implies that (in both cases)

$$d = \frac{\sigma_\varepsilon^2 - \sigma_\gamma^2}{\sigma_\varepsilon^2 + \sigma_\gamma^2},$$

so that the second-order conditions for a maximum in the two cases are, respectively:

$$\textit{Case 1:}\quad \sigma_\gamma^2 < \sigma_\varepsilon^2 \qquad \text{for } x_L > c_L + \varepsilon_L + \gamma.$$

$$\textit{Case 2:}\quad \sigma_\gamma^2 > \sigma_\varepsilon^2 \qquad \text{for } x_L < c_L + \varepsilon_L + \gamma.$$

From (13), (15), and (17), $P^L(x_L)$ will be increasing in Case 1 and decreasing in Case 2. The intuition underlying the two cases is the following. When L moves policy to the right, he triggers two conflicting effects on his reelection prospects. For a given γ^e, this moves him closer to the center of the political spectrum and increases his probability of reelection. We shall refer to this as the "[illegible]elling effect." But the shift of policy to the right also raises the forecast of γ. This moves the ideal points of all voters to the right and increases the electoral prospects of the challenging right-wing party. We refer to this as the "expectation effect." When $\sigma_\gamma^2 < \sigma_\varepsilon^2$, this shift of policy induces a moderate reduction in the reelection prospects of L via the expectation effect, and the [illegible]elling effect dominates, raising the reelection prospects of the left-wing incumbent. It therefore induces L to choose a policy that is more centrist than its ideal policy $(c_L + \varepsilon_L + \gamma)$. When $\sigma_\gamma^2 > \sigma_\varepsilon^2$, the expectation effect dominates and a shift to the right reduces, on balance, L's reelection prospects. It therefore chooses a policy that is more extreme than its ideal point.

It is shown below that in both Cases 1 and 2, a right-wing *incumbent* (RWI) is more likely to propose a relatively right-wing policy. In Case 2, as policy shifts to the right, the probability of reelection of a left-wing incumbent (LWI) falls and that of a RWI rises. Hence, policy reversals (i.e., higher chances of a right-wing policy being adopted by a left-wing incumbent) are ruled out in Case 2. Since we are interested in the conditions determining policy reversals, we concentrate on Case 1 in the following. $\sigma_\gamma^2 < \sigma_\varepsilon^2$ is, therefore, one of the conditions required for policy reversals. A relatively small σ_γ^2 means that the mass assigned to the tails of the normal distribution of γ is small in comparison to the tails of the distribution of ε_L (i.e., that the probability of γ taking extreme values is small in comparison to the probability of ε_L taking extreme values).

II. Which Party is More Likely to Implement Which Policies?

We now come to the central issue, stated in the title of this section. To do that, we focus on a comparison between the behavior of left-wing and right-wing *incumbents*.

A derivation equivalent to that of the previous section, for a right-wing incumbent (for $\sigma_\gamma^2 < \sigma_\varepsilon^2$) leads to:

$$x_R = B_R + \tfrac{1}{2}(\gamma + \varepsilon_R),$$

with the mean (equilibrium) policy choice of a right-wing incumbent being

$$(22)\quad B_R = \frac{1}{2}\left[\left(1 - \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)(c_L - h) + \left(1 + \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)(2\bar{c} - c_L)\right].$$

The probability of reelection of a right-wing incumbent as a function of his (individually optimal) policy choice is:

$$(23)\quad P^R(x_R) = \frac{1}{2(\bar{c} - \underline{c})}\,[2\bar{c} - c_L - 2\theta B_R - dx_R].$$

Notice, comparing Case 1 of (21) and (22), that as the value of office, h, increases, both parties' policies converge to the center (the Hotelling effect). On the other hand, as the degree of electoral uncertainty increases (a larger $\bar{c}$ and/or smaller $\underline{c}$), the policy of R becomes more right-wing (and that of L becomes more left-wing). Finally; c_R and c_L also enter with the expected sign.

Let $PI^i(x)$ be the probability that incumbent party i *implements* policy x, and $Q^i(x)$ be the probability that incumbent party i *proposes* policy x. For a policy to be implemented, it has to be proposed by the incumbent party *and*, given this proposal, the incumbent party has to receive public support via reelection. It follows that the probability of implementation of policy x by party i is:

$$PI^i(x) = Q^i(x)P^i(x),$$

where $P^i(x)$ are the probabilities of reelection as stated in (13) and (23).

Notice that $x_i = B_i + {}^1\!/_2(\gamma + \varepsilon_i)$ implies that

$$x_i \sim N[B_i, V],$$

with $V = (\sigma_\gamma^2 + \sigma_\varepsilon^2)/4$. Hence:

$$Q^i(x) = (2\pi V)^{-(1/2)}\exp\left\{-\frac{(x - B_i)^2}{2V}\right\}.$$

To simplify the calculations, we confine ourselves, henceforth, to the *symmetric case:* in which $\bar{c} = -\underline{c} > 0$ and $c_R = -c_L > 0$. This implies that

$$B_R = -B_L = B = \left(1 + \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)\bar{c}$$

$$- \left(1 - \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)\frac{h}{2} + \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}c_R,$$

$$P^L(x) = \frac{A + dx}{4\bar{c}}, \qquad \text{and} \qquad P^R(x) = \frac{A - dx}{4\bar{c}},$$

where

$$A = 2\left(1 - \frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)\bar{c} + \left(1 - \frac{2\sigma_\gamma^4}{\sigma_\varepsilon^2(\sigma_\varepsilon^2 + \sigma_\gamma^2)}\right) \ldots$$

$$+ \frac{\sigma_\gamma^2(\sigma_\varepsilon^2 - \sigma_\gamma^2)}{(\sigma_\varepsilon^2 + \sigma_\gamma^2)^{\cdots}}h.$$

Notice, as before, that the (average) policy ... each of the parties contains the following fa... tors: an electoral uncertainty effect (B increas... ing in $\bar{c}$), a Hotelling-Downs effect ... decreasing in h), and an ideological prefer... effect (B increasing in c_R).[13]

Let

$$\Delta PI(x) = PI^R(x) - PI^L(x).$$

When $\Delta PI(x) > 0$, policy x is more likely ... be implemented by a right-wing party than ... a left-wing party; the converse is true wh... $\Delta PI(x) < 0$.[14] Let

$$F(x) = \frac{2B}{V}x + \ln(A - dx) - \ln(A + dx).$$

LEMMA 1: *Sign*$[\Delta PI(x)] = $ *sign*$[F(x)]$.

PROPOSITION 1: *If* $Vd/B < A < 2\bar{c}$, ... *the range of policies x can be partitioned* ... *the following way:*

(1) *There is a central region* $(\underline{x}, \bar{x})$ *in wh... the conventional result obtains (poli... to the left of the center of the poli...*

[13] We assume, henceforth, that the mean policy ... posed by a RWI is larger than the mean policy pr... by a LWI. That is, $B_R > B_L$. This is the case whe... systematic difference between the ideological posi... the two parties is sufficiently large in comparison to ... (shared) love of office, h. In the symmetric case, th... plies that B_R is positive, B_L is negative, and that th... positioned at equal distances from the center of the p... ical spectrum (which is at $x = 0$).

[14] Notice that we concentrate on comparing the p... bilities of implementation attached to different ... *bents*. We believe that this is the relevant comparison ... in our (two-period) model, only the incumbent beh... an interesting strategic way; the behavior of challe... passive.

spectrum are more likely to be implemented by L, and policies to the right of the center are more likely to be implemented by R).

(2) (Policy Reversals) There is a region outside $(\underline{x}, \bar{x})$ in which very left-wing policies $(x < \underline{x})$ are more likely to be implemented by R, and very right-wing policies $(x > \bar{x})$ are more likely to be implemented by L.

(3) (Only Nixon) For even more extreme values of x, above x_R^0 and below x_L^0 (where $x_R^0 > \bar{x}$ and $x_L^0 < \underline{x}$), only the unlikely party can implement policy x (i.e., $P^R[x > x_R^0] = 0$, and $P^L[x < x_L^0] = 0$).

Proofs of Lemma 1 and Proposition 1 are provided in the Appendix. Figure 2 summarizes the intuition of the proposition. Panel A shows the probabilities of each proposal, $Q^i(x)$. Panel B shows the probabilities of winning the election as a function of the proposal, $P^i(x)$. Notice that the picture of $PI^L(x) = Q^L(x)P^L(x)$, shown in Panel C, can be obtained by increasing the mass of the right tail and decreasing the mass of the left tail of $Q^L(x)$, and conversely for R. Panel D shows the behavior of $F(x)$.

III. Comparative Statics

Figure 2, Panel C, suggests that, in general, relatively extreme policies are unlikely to be implemented by either party. The reason is that extreme realizations of the stochastic component of the mapping of policies into outcomes and of the intraparty shocks are relatively rare. However, when such events do appear, the probability that the corresponding extreme policy will be implemented by an "unlikely" party is greater than the probability that it will be implemented by the "likely" party.

Intuitively, the conditions that are conducive to policy reversals can be summarized as follows: First, the variability of intraparty policy preferences has to be large in comparison to the variability in the mapping of policies into outcomes. This assures that when the policy proposal of the incumbent party shifts towards the center, the Hotelling effect dominates the expectation effect, thus increasing the reelection prospects of the party. Second, reversals are more likely to occur the more extreme the policy that is being proposed. Since extreme policies are proposed infrequently, policy reversals will also be infrequent events that will be associated with extreme and relatively rare realizations of γ.

In order to gain additional insights, we will now perform comparative statics to see how the range of reversals, the complement of $(\underline{x}, \bar{x})$, varies with the parameters of the model. We know that $\underline{x}$, 0 and $\bar{x}$ are the three roots of $F(x) = 0$ (although only the location of $\underline{x}$ and $\bar{x}$ varies with the underlying parameters).[15] Let α denote any parameter. Applying the implicit function theorem to $F(x) = 0$,

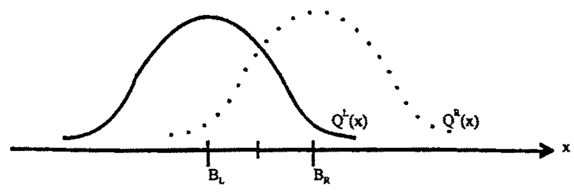
$$(24)\quad \frac{dx}{d\alpha} = -\frac{1}{F'(x)}\left\{\frac{2x}{V^2}\left(\frac{\partial B}{\partial \alpha}V - \frac{\partial V}{\partial \alpha}B\right) + \frac{\dfrac{\partial A}{\partial \alpha} - x\dfrac{\partial d}{\partial \alpha}}{A - dx} - \frac{\dfrac{\partial A}{\partial \alpha} + x\dfrac{\partial d}{\partial \alpha}}{A + dx}\right\},$$

where $x = \underline{x}, 0, \bar{x}$.

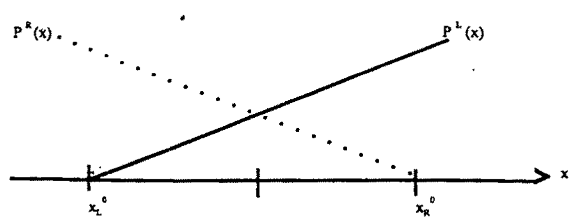
Applying (24) to the electoral uncertainty parameter $\bar{c}$, we obtain

$$(25)\quad \frac{dx}{d\bar{c}} = -\frac{x}{F'(x)}\left\{\frac{2\left(1 + \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)}{V} + \frac{4\left(1 - \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)d}{A^2 - (dx)^2}\right\}.$$

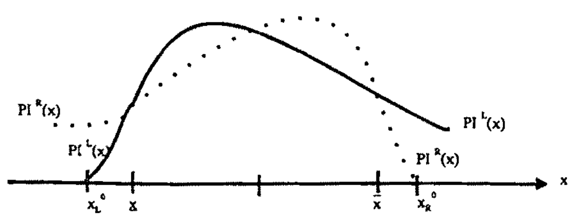
[15] Note that the probability of implementation of the policy $x = 0$, at the center, is identical for both parties. The fact that $F(0) = 0$ is a consequence of the postulated symmetry, but the general qualitative nature of Proposition 1 does not depend on the symmetry assumption.



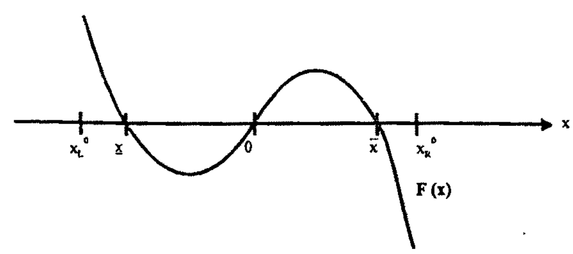
A. Probabilities of Policy Proposals by Incumbents L and R



B. Probabilities of Reelection as a Function of Policy Proposals



C. Probabilities of Implementation: $PI_i(x) = Q_i(x)P_i(x)$



D. $F(x) = PI^R(x) - PI^L(x)$

Figure 2. Probability of Policy Implementation by Each Type of Incumbent

Notice that

$$-\frac{x}{F'(x)}\begin{cases} <0 & \text{at } x = \underline{x} < 0 \\ =0 & \text{at } x = 0 \\ >0 & \text{at } x = \bar{x} > 0 \end{cases}$$

and that the expression in braces in (25) is positive. This leads to:

RESULT 1: *An increase in electoral uncertainty (measured by $\bar{c}$), reduces the range of "policy reversals."*

Similarly,

$$\frac{dx}{dc_R} = -\frac{2x}{F'(x)}\left\{\frac{\sigma_\gamma^2}{V\sigma_\varepsilon^2} + \frac{d\left(1 - 2\frac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\theta\right)}{A^2 - (dx)^2}\right\}$$

Since $(\sigma_\gamma^2/\sigma_\varepsilon^2) < 1$, $(1 - 2(\sigma_\gamma^2/\sigma_\varepsilon^2)\theta)$ is positive. It is easy to see the following:

RESULT 2: *An increase in the degree of party polarization or ideological distance between the parties (measured by c_R), decreases the range of "policy reversals."*[16]

Following a similar procedure, we are able to show:

RESULT 3: *An increase in uncertainty about the incumbent's ideological position (measured by σ_ε^2), increases the range of "policy reversals."*

Result 3 suggests that the policy reversals we are studying should be more common in countries characterized by "catchall" parties that comprise a wide spectrum of relatively heterogeneous constituencies, or by coalition governments within which there are agreements (deals) which are not easily observed by the general public. The Peronist movement in Argentina fits the first interpretation quite well and some of the broad "unity" governments in Israel conform to the second.

IV. Implications for Credibility[17]

Politicians usually justify their policy proposals by claiming that the state of the wor

[16] Schultz (1996) addresses other effects of polarization in a model that also emphasizes asymmetric information about a parameter that he calls "the functioning of the economy," which plays the exact same role as our
[17] The notions of credibility that we use in this are by no means the only ones possible. Alterna

is the one that (if believed) would elicit maximum public support for their proposed policies. In order to elicit support for their policies, politicians argue that those policies are beneficial for the majority of the population. In the context of our model they would argue that γ is quite large (small) when they propose a right (left)-wing policy.[18]

In this section, we use two notions which measure the credibility of (implicit) statements about γ by looking at the posterior beliefs about γ held by voters after having observed policy proposals. We shall continue to refer to any policy that is to the right of the center of the political spectrum as "right wing" and to any policy that is to the left of this center as "left wing." (For the symmetric case, the center of the political spectrum is at 0.)

1. The credibility of a given policy maker when he proposes a relatively left (right)-wing policy can be characterized as the posterior probability that γ is smaller (larger) than a given value of γ.

The public's observation of the policy proposal x, given its knowledge of B_i and of $b_i = \frac{1}{2}$, is equivalent to an observation of $(\gamma + \varepsilon_i) \equiv 2(x - B_i)$, for $i = L, R$. We are interested in inferences about the unknown value of γ made from the observation of $(\gamma + \varepsilon_i)$. This observation is a random drawing from a normal distribution with an unknown mean γ and variance σ_ε^2. The prior distribution over that unknown mean is $N(0, \sigma_\gamma^2)$. Using Theorem 1 of Section 9.5 of De Groot (1970 p. 167), the posterior distribution of γ, $f_i(\gamma, x)$ is $N(\gamma_i^e(x), \Sigma^2)$, where $\gamma_i^e(x) = \theta(\gamma + \varepsilon_i) = 2\theta(x - B_i)$, and $\Sigma^2 = \sigma_\varepsilon^2 \sigma_\gamma^2 / (\sigma_\varepsilon^2 + \sigma_\gamma^2)$.

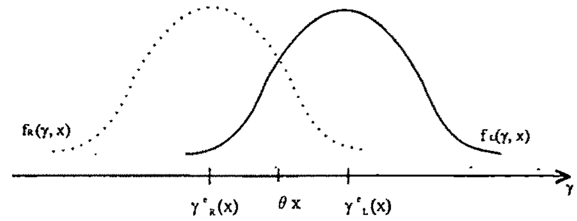


FIGURE 3. POSTERIOR BELIEFS ABOUT γ, AFTER POLICY ANNOUNCEMENT x

Recalling that $B_L < B_R$, it is easy to see that $\gamma_R^e(x) < \gamma_L^e(x)$ for all x. This means that any given policy, x, is interpreted by the public as being associated with a lower (posterior mean) value of γ when this policy is proposed by a RWI than when it is proposed by a LWI. The intuition is simple. Being aware of the systematic ideological differences between the two parties, the public expects to observe the *same* policy from both incumbents only when the RWI observes a lower value of γ than his leftist counterpart.

The posterior distribution that the public assigns to γ after observing a proposal x, is depicted (for the symmetric case) in Figure 3. It is clear that, for any given x,

$$(26) \quad f_L(\gamma, x) < f_R(\gamma, x) \qquad \text{for } \gamma < \theta x,$$

$$f_L(\gamma, x) > f_R(\gamma, x) \qquad \text{for } \gamma > \theta x$$

and

$$F_L(\gamma, x) < F_R(\gamma, x), \ \forall \gamma,$$

where F is the cumulative distribution function of f.

RESULT 4: *(Credibility of implicit statements about γ across policy makers) For any given (common) policy proposal: (i) a left-wing policy maker is more credible than a right-wing policy maker when he claims that right-wing policies are desirable, and (ii) a right-wing policy maker is more credible than a left-wing policy maker when he claims that left-wing policies are desirable.*

2. Consider now "extreme" policies. Extreme policies are defined as policies that are

[17] Issues of credibility are discussed in Chapter 11 of Cukierman (1992). See also the discussion of Allan Drazen and Paul Masson (1994) in Section V of this paper.

[18] In Roemer (1994), political parties make announcements about "the way the economy works" (γ in our model), and then propose policies. In his model, there is complete convergence of policies (the median voter theorem) but different parties announce different γ's in an attempt to influence people's reduced-form preferences. Roemer does not model the formation of beliefs explicitly. Instead he postulates a mapping of announcements into beliefs. Here, we deduce the belief-formation process from rational (Bayesian) behavior of voters.

to the right of the average policy proposal made by a RWI or to the left of the average policy proposal made by a LWI. Politicians can convince a majority that an extreme policy is necessary by claiming that there has been an extreme realization of γ. We therefore measure the credibility of policy proposals by the posterior probability assigned by the public to the event "γ is larger than a sufficiently large and positive $\hat{\gamma}$" when an extreme right-wing policy x is proposed—and by the posterior probability assigned by the public to the event "γ is smaller than $-\hat{\gamma}$" when an extreme left-wing policy is proposed.

Consider a policy $x > B > 0$ and its mirror image (in the symmetric case) $-x$, and let $\hat{\gamma} > 2\theta(B + x)$. We know that $\gamma_L^e(x) = 2\theta(B + x)$ and that $\gamma_L^e(-x) = 2\theta(B - x)$. It is easy to see that $B > 0$ implies

$$\gamma_L^e(-x) - (-\hat{\gamma}) > \hat{\gamma} - \gamma_L^e(x);$$

that is, $\hat{\gamma}$ deviates from $\gamma_L^e(x)$ less than $(-\hat{\gamma})$ deviates from $\gamma_L^e(-x)$. This yields

$$(27) \qquad [1 - F_L(\hat{\gamma}, x)] > F_L(-\hat{\gamma}; -x).$$

A similar derivation for a right-wing incumbent yields:

$$(28) \qquad F_R(-\hat{\gamma}, -x) > [1 - F_R(\hat{\gamma}, x)].$$

In words, (27) and (28) imply:

RESULT 5: *(Credibility of a given policy maker across extreme policies) (i) A left-wing policy maker has more credibility when he proposes a significant policy shift to the right than when he proposes a significant policy shift to the left. (ii) A right-wing policy maker has more credibility when he proposes a significant policy shift to the left than when he proposes a significant policy shift to the right.*

V. Conclusions

When governments have better information than do voters about the way in which policies map into outcomes, policy proposals con[...] information. This may lead to situations in which extreme, but rarely proposed, policies are more likely to be implemented by "un-likely" actors. A necessary, but not sufficient condition for such reversals is that the uncertainty about the preferred policy position of parties be sufficiently large compared to the uncertainty about the state of the world.

A main result of the paper is that policy reversals are more likely following the realization of extreme and relatively unlikely values of parameters that map policy choices into outcomes. A corollary to this result is that policy reversals occur infrequently.

This logic can be applied to several policy arenas. In Cukierman and Tommasi (1998) we apply it to the specific issue of market-oriented reforms undertaken by many countries over the last decade. We argue there that politicians coming from the left of the spectrum, when faced with the fact that such policies were necessary, had a comparative advantage in convincing people of the long-run necessity of these changes, even if they hurt now. There is some preliminary evidence in favor of this point. John Williamson asked the contributors to his 1994 edited volume to check the conventional wisdom that market-oriented reforms are creatures of right-wing governments. Little support was found for such an association, as summarized by Williamson and Stephan Haggard (1994). As a matter of fact, in only three of their 13 cases was market-oriented reform implemented by what they classified as right-wing governments. Interestingly, these three cases included the two military dictatorships in their sample (Chile and Korea). This seems consistent with the prediction that, under democratic conditions, large shifts in policy are more likely to involve reversals of a party's traditional policy position.[19] In Cukierman and Tommasi (1998) we

[19] On a related issue, Cesar Martinelli and Tommasi (1997) argue that the implementation of reform packages might suffer from time-consistency problems. Groups that benefit from early reforms but suffer from later ones might blockade the later stages, making some reform paths time inconsistent. An implication of the logic of this paper is that policy makers may, then, initiate a reform sequence by implementing those measures that hurt their own constituencies.

discuss some cases, other than those included in the Williamson volume, which display similar characteristics.

Some well-known foreign policy reversals are also consistent with the framework of this paper, such as the opening up to the People's Republic of China by staunch anti-Communist Richard Nixon, and the trading of land for peace by hawkish Israeli Prime Minister Begin, cited above. Begin managed to overcome strong, and sometimes violent, opposition to the dismantling of Jewish settlements in the Sinai Peninsula largely because he had a long and respectable record against such a policy. His record as a hawk helped him convince voters that the agreement was beneficial for a majority of Israelis. If the same policy had been adopted by the dovish Labor Party, mobilization of sufficient public support would have been harder, if not impossible. Similarly, if the opening up to China had been attempted by Humphrey rather than by Nixon, opposition to that policy would have been more difficult to overcome.

The analysis has some interesting implications for the question of credibility. In particular, we show that the credibility of a policy depends on the ideological identity of the policy maker proposing it, as well as on the policy he proposes. This is related, although not identical, to the distinction between the credibility of policies and the credibility of policy makers made by Drazen and Masson (1994).[20] In both papers, the notion of credibility employed depends upon the policy-policy maker pair. But there are also important differences. In Drazen and Masson credibility refers to the likelihood that an announced policy will be carried out (under discretion), within a framework in which there is uncertainty about the policy maker's type and about economic shocks that may alter his *ex post* preferred policy. Here there is no uncertainty about the policy of the incumbent since he is committed to the announced policy. Hence, our concept of credibility refers to the incumbent's ability to convince voters that the policy he will implement, if reelected, is better from the point of view of the majority than the challenging party's policy. In both cases there is uncertainty about the policy that will, ultimately, be implemented. In our case the uncertainty is due to political competition in conjunction with electoral uncertainty, while in Drazen and Masson it is due to uncertainty about the policy maker's type in conjunction with uncertainty about economic shocks. Each notion of credibility is natural whitin the framework under consideration. Combining the two frameworks may yield more general insights about basic institutional determinants of credibility. We leave this task for future work.

[20] See also Drazen (1996) for a more general treatment.

APPENDIX

Note: This Appendix focuses on the case $\sigma_\gamma^2 < \sigma_\varepsilon^2$, which is necessary for policy reversals.

1. Conditions on Exogenous Parameters for Assumption 1

Assumption 1 requires that

$$\text{(A1)} \quad x_L = B_L + \tfrac{1}{2}(\gamma + \varepsilon_L) < c_R + \theta(\gamma + \varepsilon_L) = x_R^e.$$

Using (21) we can rearrange (A1) as:

$$\text{(A2)} \quad \frac{(\gamma + \varepsilon_L)}{(\sigma_\varepsilon^2 + \sigma_\gamma^2)} < \frac{2\left(1 + \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)(c_R - \underline{c}) - \left(1 - \dfrac{\sigma_\gamma^2}{\sigma_\varepsilon^2}\right)h}{(\sigma_\varepsilon^2 - \sigma_\gamma^2)}.$$

The left-hand side of (A2) is a standard normal. Therefore, the condition is satisfied with very high probability if c_R is sufficiently larger than $\underline{c}$ and/or the (positive) difference $(\sigma_\varepsilon^2 - \sigma_\gamma^2)$ is sufficiently small. (It is also more likely to be satisfied, the smaller the value of h.)

2. Taking Care of Corners

Note: For the sake of brevity, we work directly with the symmetric case, as in Section II.

The analysis in the text was undertaken under the implicit assumption that the probability of reelection described by equation (13) belongs to (0, 1). For the incumbent L, this means that

$$\text{(A3)} \qquad P^L(x) = \frac{A + dx}{4\bar{c}} \in (0, 1).$$

Let $\mu_L = \gamma + \varepsilon_L$. From the policy function $x_L = B_L + {}^1\!/_2\mu_L$, it is easy to see that (A3) requires

$$\mu_L \in (\mu_L^0, \mu_L^1),$$

where $\mu_L^0 = -2(A/d + B_L)$ and $\mu_L^1 = 2[(4\bar{c} - A)/d - B_L]$. (The corresponding values of x_L are $x_L^0 = -A/d$ and $x_L^1 = (4\bar{c} - A)/d$.)

In equilibrium, the probability of reelection $P^L(x)$ can be thought of as the stochastic aggregation of the median voter's best response to the policy function $x_L(\mu_L)$. This policy function is, in turn, the best response to $P^L(x)$. We have established in Section I that $P^L(x) = (A + dx)/4\bar{c}$ and $x_L(\mu_L) = B_L + {}^1\!/_2\mu_L$ are best responses to each other, which is true only for $\mu_L \in (\mu_L^0, \mu_L^1)$. Now we proceed to describe behavior outside that range.

(i) When $\mu_L < \mu_L^0$, $P^L(x) = 0$ and $x_L(\mu_L) = B_L + {}^1\!/_2\mu_L$ are best responses to each other. For $x_L < -A/d$, all potential median voters prefer to vote for the challenger R, so that $P^L(x) = 0$. Therefore, the incumbent faces the decision of either definitely losing the election, or committing to a policy $x_L > -A/d$. The first course of action gives a payoff of $-(c_R - c_L) = -2c_R$ and the second, an expected payoff of

$$\left(\frac{A + dx}{4\bar{c}}\right)[h - x_L + c_L + \mu_L]$$
$$- \left(1 - \frac{A + dx}{4\bar{c}}\right)2c_R.$$

It can be shown that $[h - x_L + c_L + \mu_L] < -2c_R$ for $x_L > -A/d$ and $\mu_L < \mu_L^0$, implying that the first course of action is preferred. This is demonstrated by showing that this inequality is satisfied for $x_L = -A/d$ and $\mu_L = \mu_L^0$ and by noting that it is a fortiori satisfied for hi[gher] values of x_L and lower values of μ_L.

Hence, when $\mu_L < \mu_L^0$, the incumbent cho[oses] a policy that gives $P^L(x) = 0$ since, in orde[r to] attain $P^L(x) > 0$ he would have to commit [to a] policy too far from his (*ex post*) preferred [one]. This means that any policy function $x_L(\mu_L)$ [that] gives $x_L < -A/d$ $\forall \mu_L < \mu_L^0$ is consistent [with] $P^L(x) = 0$ $\forall x < -A/d$. In particular, $x_L(\mu_L) = B_L + {}^1\!/_2\mu_L$ is one such function.

(ii) When $\mu_L > \mu_L^1$, the previous poli[cy] function induces an x_L so large that $(A + dx)/4\bar{c} > 1$, which means that all potential media[n] voters would vote for the incumbent (so the $P^L(x)$, properly defined, equals 1). In th[is] range, the electoral effect on policy becom[es] irrelevant, since the incumbent is going to be reelected with certainty. Hence, the incumbe[nt] party adopts its most preferred policy, $x_L = c_L + \mu_L$. If the public knows that $x_L = c_L + \mu_L$, [it] forms expectations according to $\gamma^e = \theta(x_L - c_L)$. This expectation function leads to the in[-]cumbent L being reelected with probabil[ity] one for $x_L \geq 2\bar{c}/(1 - \theta) - c_R$ or, equivalent[ly], $\mu_L > 2\bar{c}/(1 - \theta)$. Notice that, depending on the values of the underlying parameters, μ_L^1 can [be] smaller than, larger than, or equal to $2\bar{c}/(1 - \theta)$. For the sake of brevity, assume that $\mu_L^1 = 2\bar{c}/(1 - \theta)$. In that case, $\{P^L(x) = (A + dx)/4\bar{c},\ x_L(\mu_L) = B_L + {}^1\!/_2\mu_L\}$ for $\mu_L < \mu_L^1$, $\{P^L(x) = 1,\ x_L(\mu_L) = c_L + \mu_L\}$ for $\mu_L > \mu_L^1$, and both [are] best-response pairs for $\mu_L \in (2\bar{c}/(1 - \theta), \mu_L^1)$. If, as a refinement, we require continuity [of] the $P^L(x)$ function, the full characterization [of] equilibrium is:

$$x_L(\mu_L) = \begin{cases} B_L + \frac{1}{2}\mu_L & \text{for } \mu_L \leq \mu_L^1 \\ c_L + \mu_L & \text{for } \mu_L > \mu_L^1 \end{cases}$$

and

$$P^L(x) = \begin{cases} 0 & \text{for } x \leq -A/d \\ \dfrac{A + dx}{4\bar{c}} & \text{for } -A/d < x < (4\bar{c} - A)/d \\ 1 & \text{for } x \geq (4\bar{c} - A)/d \end{cases}$$

3. PROOF OF LEMMA 1:

$$\Delta PI(x) = \frac{(2\pi V)^{-(1/2)}}{4\bar{c}}\left\{(A-dx)\exp\left[-\frac{(x-B)^2}{2V}\right] - (A+dx)\exp\left[-\frac{(x+B)^2}{2V}\right]\right\}$$

$$= \frac{(2\pi V)^{-(1/2)}}{4\bar{c}}\exp\left[-\frac{x^2+B^2}{2V}\right] \times \left\{(A-dx)\exp\left[\frac{B}{V}x\right] - (A+dx)\exp\left[-\frac{B}{V}x\right]\right\}.$$

Clearly,

$$\Delta PI(x) \gtreqless 0 \quad \text{as} \quad (A-dx)\exp\left[\frac{B}{V}x\right] \gtreqless (A+dx)\exp\left[-\frac{B}{V}x\right].$$

Taking the logarithms (a monotonic transformation) of both sides of the inequality, we arrive at

$$\Delta PI(x) \gtreqless 0 \quad \text{as} \quad \frac{B}{V}x + \ln(A-dx) \gtreqless -\frac{B}{V}x + \ln(A+dx),$$

as claimed.

4. PROOF OF PROPOSITION 1:

First notice that Panels C and D of Figure 2 are equivalent.

In order for the curves to be as presented, it is necessary that the interior regions (x_L^0, x_L^1) and (x_R^0, x_R^1) overlap, or $x_L^1 = (4\bar{c} - A)/d > (A - 4\bar{c})/d = x_R^1$, which is to say $4\bar{c} > A$.[21]

We now proceed to characterize $F(x)$. Note that $F(0) = 0$, that $F(x) \to \infty$ as $x \to x_L^0$ from above, and that $F(x) \to -\infty$ as $x \to x_R^0$ from below. Hence, we know from continuity, that $F(x) = 0$ at least three times, at $x = \underline{x} \in (x_L^0, 0)$, at $x = 0$, and at $x = \bar{x} \in (0, x_R^0)$.

We can verify that these are indeed the three crossings by analyzing

$$F'(x) = \frac{2B}{V} - \frac{d}{(A-dx)} - \frac{d}{(A+dx)}.$$

Notice that $F'(x) = 0$ at

$$\text{(A4)} \qquad x = \frac{1}{d}\sqrt{1 - \frac{Vd}{AB}}.$$

Since $Vd < AB$, $F'(x) = 0$ has exactly two real roots, one positive and one negative, which are equidistant from zero.

Also,

$$F''(x) = -d^2\left[\frac{1}{(A-dx)^2} - \frac{1}{(A+dx)^2}\right].$$

It is easy to see that $F''(x) < 0$ for $x > 0$ and that $F''(x) > 0$ for $x < 0$. This implies that the negative root of (A4) corresponds to a minimum of $F(x)$, and that the positive root of (A4) corresponds to a maximum of $F(x)$. Hence, in the range (x_L^0, x_R^0), the slope of $F(x)$ switches signs three times. Since $F(x)$ is positive near the lower end of this range, negative near its upper end, and 0 at $x = 0$, it follows that $F(x)$ has exactly three roots and that Figure 2, Panel D, correctly represents the curve. Since the figure and the proposition are equivalent, this completes the proof.

[21] For the detailed drawing of Figure 2, it matters whether x_L^1 is smaller than, equal to, or greater than x_R^0, the value of x such that $P^R(x_R^0) = 0$. The figure is drawn with [illegible], which implies $2\bar{c} > A$, which is stronger than the [illegible] in the text.

REFERENCES

Alesina, Alberto and Cukierman, Alex. "The Politics of Ambiguity." *Quarterly Journal of Economics*, November 1990, *105*(4), pp. 829–50.

Austen-Smith, David. "Information Transmission in Debate." *American Journal of Political Science*, February 1990, *34*(1), pp. 124–52.

———. "Strategic Models of Talk in Political Decision Making." *International Political Science Review*, January 1992, *13*(1), pp. 45–58.

———. "Information Acquisition and Orthogonal Argument," in William Barnett, Melvin Hinich, and Norman Schofield, eds., *Political economy: Institutions, competition and representation*. Cambridge: Cambridge University Press, 1993, pp. 407–36.

Berndhart, M. Daniel and Ingberman, Daniel. "Candidate Reputations and the Incumbency Effect." *Journal of Public Economics*, June 1985, *27*(1), pp. 47–67.

Calvert, Randall. *Models of imperfect information in politics*, Vol. 6. Chur, Switzerland: Harwood Academic Publishers, 1986.

Coughlin, J. Peter. *Probabilistic voting theory*. Cambridge: Cambridge University Press, 1992.

Crawford, Vincent and Sobel, Joel. "Strategic Information Transmission." *Econometrica*, November 1982, *50*(6), pp. 1431–51.

Cukierman, Alex. *Central bank strategy, credibility, and independence: Theory and evidence*. Cambridge, MA: MIT Press, 1992.

Cukierman, Alex and Liviatan, Nissan. "Optimal Accommodation by Strong Policymakers under Incomplete Information." *Journal of Monetary Economics*, February 1991, *27*(1), pp. 99–127.

Cukierman, Alex and Tommasi, Mariano. "Credibility of Policymakers and of Economic Reform," in Federico Sturzenegger and Mariano Tommasi, eds., *The political economy of economic reforms*. Cambridge, MA: MIT Press, 1998 (forthcoming).

De Groot, Morris. *Optimal statistical decisions*. New York: McGraw-Hill, 1970.

Downs, Anthony. *An economic theory of democracy*. New York: Harper and Row, 1957.

Drazen, Allan. "Policy Signalling in the Open Economy: A Re-examination." Working Paper No. 19, Center for International Economics, University of Maryland, March 1996.

Drazen, Allan and Masson, Paul. "Credibility of Policies versus Credibility of Policymakers." *Quarterly Journal of Economics*, August 1994, *109*(3), pp. 735–54.

Gilligan, Thomas. "Information and the Allocation of Legislative Authority." *Journal of Institutional and Theoretical Economics*, March 1993, *149*(1), pp. 321–41.

Gilligan, Thomas and Krehbiel, Keith. "Asymmetric Information and Legislative Rules with a Heterogeneous Committee." *American Journal of Political Science*, May 1989, *33*(2), pp. 459–90.

Harrington, Joseph E., Jr. "Economic Policy, Economic Performance, and Elections." *American Economic Review*, March 1993, *83*(1), pp. 27–42.

Ingberman, Daniel. "Campaign Contributions, Candidate Reputations, and Spatial Competition." Carnegie Mellon Graduate School of Industrial Administration Working Paper Series: WP #6-84-85, November 1984.

Kircheimer, Otto. "The Transformation of the Western European Party Systems," in Joseph La Palombra and Myron Weiner, eds., *Political parties and political development*. Princeton, NJ: Princeton University Press, 1966, pp. 177–200.

Laver, Michael and Schofield, Norman. *Multiparty government: The politics of coalition in Europe*. New York: Oxford University Press, 1990.

Lupia, Arthur. "Busy Voters, Agenda Control, and the Power of Information." *American Political Science Review*, June 1992, *86*(2), pp. 390–403.

Martinelli, Cesar and Tommasi, Mariano. "Sequencing of Economic Reforms in the Presence of Political Constraints." *Economics and Politics*, July 1997, *9*(2), pp. 115–31.

Matsusaka, John. "The Economic Approach to Democracy," in Mariano Tommasi and Kathryn Ierulli, eds., *The new economics of human behavior*. Cambridge: Cambridge University Press, 1995, pp. 140–54.

Page, Benjamin and Shapiro, Robert. "Effects of Public Opinion on Policy." *American Political Science Review*, March 1983, *77*(1), pp. 175–90.

Rodrik, Dani. "The Positive Economics of Policy Reform." *American Economic Review*, May 1993 (*Papers and Proceedings*), *83*(2), pp. 356–61.

Roemer, John. "The Strategic Role of Party Ideology When Voters Are Uncertain

How the Economy Works." *American Political Science Review*, June 1994, *88*(2), pp. 327–35.

Rogoff, Kenneth. "Equilibrium Political Budget Cycles." *American Economic Review*, March 1990, *80*(1), pp. 21–36.

Rogoff, Kenneth and Sibert, Anne. "Elections and Macroeconomic Policy Cycles." *Review of Economic Studies*, January 1988, *55*(1), pp. 1–16.

Schultz, Christian. "Polarization and Inefficient Policies." *Review of Economic Studies*, April 1996, *63*(2), pp. 331–44.

Williamson, John. *The political economy of policy reform.* Washington, DC: Institute for International Economics, 1994.

Williamson, John and Haggard, Stephan. "The Political Conditions for Economic Reform," in John Williamson, ed., *The political economy of policy reform.* Washington, DC: Institute for International Economics, 1994, pp. 527–36.

What Price Coordination? The Efficiency-Enhancing Effect of Auctioning the Right to Play

By Vincent Crawford and Bruno Broseta*

A model is proposed to explain the results of recent experiments in which subjects repeatedly played a coordination game, with the right to play auctioned each period in a larger group. Subjects bid the market-clearing price to a level recoverable only in the efficient equilibrium and then converged to that equilibrium, although subjects playing the game without auctions converged to inefficient equilibria. The efficiency-enhancing effect of auctions is reminiscent of forward induction, but is not explained by equilibrium refinements. The model explains it by showing how strategic uncertainty interacts with history-dependent learning dynamics to determine equilibrium selection. (*JEL* C73, C92, C51)

Coordination is central to many questions in economics, from the determination of bargaining outcomes to the design of incentive schemes, the efficacy of implicit contracts, and the influence of expectations in macroeconomics. Such questions are usually modeled as noncooperative games with multiple Nash equilibria, and analyzed under the assumption that players can realize any desired equilibrium. Yet this begs the questions of whether, and how, coordination comes about and how the environment influences equilibrium selection—questions that lie at the heart of most applications.

Although recent advances in game theory have added much to our understanding of coordination, there remains a large gap between theory and experience that is unlikely to be closed by theory alone. Further progress seems likely to depend on combining theory with empirical evidence on strategic behavior. Experiments are a particularly useful source of such evidence, in part because they make it possible to observe the entire coordination process. Crawford (1997) surveys a number of recent studies in which subjects repeatedly played coordination games, uncertain only about each other's strategy choices. The typical result was convergence to an equilibrium, often with a systematic pattern of equilibrium selection in the limit. Explaining such patterns promises to shed considerable light on coordination, in the field as well as the laboratory.

In this paper we seek to explain some of the most intriguing evidence on coordination we have seen, from the experiment of John Van Huyck et al. (henceforth "VHBB") (1993).[1] Their subjects repeatedly played a nine-person coordination game with seven symmetric, Pareto-ranked equilibria from the experiment of VHBB (1991), with subjects' payoffs and best responses determined by their own actions, called *efforts*, and an order statistic of all subjects' efforts, in this case the median

* Crawford: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093; Broseta: Department of Economics, University of Arizona, McClelland Hall, Room 401, Tucson, AZ 85721. We thank Raymond Battalio, Andreas Blume, Miguel Costa Gomes, Yong-Gwan Kim, Mark Machina, Daniel McFadden, Ronald Oaxaca, Alvin Roth, Jason Shachat, Joel Sobel, Glenn Sueyoshi, John Van Huyck, and two referees for helpful comments; Miguel Costa Gomes and Mahesh Johari for research assistance; Raymond Battalio, Richard Beil, Gerard Cachon, Colin Camerer, and John Van Huyck for sharing their experimental data; Raymond Battalio, John Van Huyck, and Roberto Weber for conducting additional experiments for us; and the National Science Foundation, the Centro de Formación del Banco de España, and the College of Business and Public Administration, University of Arizona, for research support.

[1] See also VHBB's (1990) experiments, where the order statistic was the minimum.

However, instead of endowing nine subjects with the right to play the game as in their 1991 experiments, VHBB auctioned its nine positions each period in a group of 18. The structure was publicly announced at the start; the market-clearing price and median effort were publicly announced each period; and explicit communication was prohibited throughout.

VHBB's (1993) design is of unusual economic interest. Its 1991 precursor, in which nine subjects played the same median coordination game without auctions, is among the simplest models of the emergence of conventions to solve coordination problems, with a range of possible outcomes and a natural measure of their efficiency. The median game is similar in structure to several important models, from the Stag Hunt example Rousseau used to motivate his analysis of the social contract to Keynes's beauty contest analogy and the more prosaic macroeconomic coordination models surveyed in Russell Cooper and Andrew John (1988). Its efficient equilibrium is plainly the "correct" coordinating principle, but it is best for a subject to play his part of that equilibrium only if he thinks it likely that enough other subjects will do so. As in the other coordination games VHBB (1990, 1991) studied, there is a tension between the higher payoff of the efficient equilibrium and the greater robustness of other equilibria to subjects' uncertainty about each other's responses, which we call *strategic uncertainty*. Finally, auctioning the right to play is an interesting form of preplay communication, in which subjects' willingness to pay may signal how they expect to play, and thereby alleviate the tension due to strategic uncertainty.[2] The auctions also capture important aspects of "general equilibrium" analogs of VHBB's earlier environments, in which players choose among coordination games with the market-clearing price analogous to the opportunity-costs determined by their best alternatives.

Auctioning the right to play the median game had remarkable consequences. When that game was played without auctions in VHBB's (1991) experiment, most subjects initially chose inefficiently low efforts, and six out of six subject groups converged to inefficient equilibria. Auctions might be expected to yield more efficient outcomes simply because subjects have diverse beliefs about each other's efforts, auctions select the most optimistic subjects to play, and optimism favors efficiency in the median game. But VHBB's (1993) subjects did much better than this argument suggests: In eight out of eight groups, they bid the market-clearing price to a level recoverable only in the efficient equilibrium and then converged to that equilibrium.[3] Convergence was very fast, essentially complete within three to five periods. The limiting outcome was consistent with subgame-perfect equilibrium in the *stage game* consisting of the auction followed by the coordination game, and subjects' behavior suggests that their beliefs were focused as in the intuition for forward induction refinements, in which players infer from their partners' willingness to pay to play a game that they expect their payoffs to repay their costs, and intend to play accordingly (Elchanen Ben-Porath and Eddie Dekel, 1992).[4]

The efficiency-enhancing effect of auctioning the right to play in VHBB's (1993) experiment suggests a novel and potentially important way in which competition might promote efficiency. Although it conveys a powerful impression by itself, it raises as many questions as it answers. Was the outcome

[2] The auction is an unusual form of preplay communication in that players' messages can directly influence their payoffs, hence are not "cheap talk"; and they are communicated only through an aggregate, the market-clearing price.

[3] Roberto Weber (1994) has replicated this result in a closely related environment, and Gerard Cachon and Colin Camerer (1996) have verified its robustness and refined its interpretation. Their experimental designs and results, and the extension of our analysis to explain them, are discussed in Crawford and Broseta (1995).

[4] The intuition for forward induction seems to favor the efficient equilibrium, and subjects' limiting behavior was consistent with forward induction, as it is usually formalized, as well as subgame perfectness. But both refinements are consistent with any of the seven symmetric pure-strategy equilibria in the coordination game (with full surplus extraction in the auction), so neither helps to explain VHBB's result.

VHBB observed inevitable in their environment? How would the strength of the effect vary with the *treatment variables*—the number of players in the auction and the coordination game and the order statistic that determines the robustness of the efficient equilibrium—which never varied in the experiment? Would the effect extend to environments beyond those that directly generalize VHBB's designs?

In principle such questions could be answered piecemeal by further experiments, but only a theory that elucidates the mechanism behind VHBB's result can provide a firm basis for generalization and realize its full power to inform analysis. Such a theory can change the way we think about many applications in which the participants in coordination games are determined by a sorting process like VHBB's auction or its general equilibrium analogs.

The models in Cooper and John (1988), for instance, view the entire economy as a coordination game, but it may be more realistic to view it as composed of sectors, regions, or firms, each of which is a coordination game. Because participants must often choose among these, the economy may be closer to VHBB's auction environment than the games with exogenous participation of their earlier experiments. If so, VHBB's result suggests that modeling the economy as a single coordination game systematically overstates the power of coordination failure to explain underemployment. Alternatively, consider a coordination model of Hong Kong's economy that takes into account the high costs most inhabitants paid to get there.[5] Did Hong Kong do so well by attracting entrants who expected efficient coordination, as in our "optimistic subjects" effect; via the power of entry barriers to focus beliefs on better coordination outcomes, as in our "forward induction" effect; by having rules that favor efficient coordination, perhaps as in the "robustness" effect discussed below; or simply by exploiting its natural advantages and "skimming the cream" from mainland China? Only a theory that distinguishes and quantifies these effects can give convincing answers to such questions.

In this paper we propose a theory to explain VHBB's (1993) result, generalizing our analyses of VHBB's (1990, 1991) results in Crawford (1995) and Broseta (1993a, 1993b) to a class of coordination games with auctions in which VHBB's treatment variables can take arbitrary values.[6] The main difficulty our analysis must address is explaining equilibrium selection; the rapid convergence to equilibrium in the stage game in VHBB's (1993) experiment and most of their earlier treatments is easily explained by learning models.

In VHBB's (1990, 1991) experiments, the dynamics and limiting outcomes varied systematically with the size of the groups playing the game and the order statistic that determined their payoffs and best responses, with strong downward drift and convergence to the equilibrium with lowest effort in VHBB's (1990) large-group minimum games, but no drift and consistent "lock-in" on the initial median in VHBB's (1991) median games. These and other variations in equilibrium selection across VHBB's treatments discriminate sharply among alternative theories of strategic behavior, both traditional and adaptive.

Our analyses explained these variations as persistent effects of interactions between strategic uncertainty and history-dependent learning dynamics. One would expect the level of strategic uncertainty to decline gradually to zero as players learn to forecast each others' responses. Unless this decline is very slow, the learning dynamics lock in on a particular equilibrium in the limit, and the model's implications for equilibrium selection can be summarized by the prior probability distribution of that equilibrium, which is normally nondegenerate due to the persistent effects of strategic uncertainty. The distribution of the

[5] We are grateful to a co-editor for suggesting this illustration, which is closer to VHBB's auction environment because Hong Kong's position has been nearly unique.

[6] Yong-Gwan Kim (1996) suggests interesting alternative explanations of VHBB's result, based on adaptive dynamics and axioms in the spirit of evolutionary stability.

limiting equilibrium is determined by players' learning rules, the treatment variables, and the cumulative effects of strategic uncertainty on the dynamics. The key to our explanation is that strategic uncertainty generates what we shall call a robustness effect, which gives the learning dynamics a negative, zero, or positive drift for order statistics respectively below, at, or above the median. The magnitude of the drift increases with the level of strategic uncertainty, and for sufficiently large initial levels, declining over time, it makes the prior probability distribution of the limiting equilibrium vary across VHBB's (1990, 1991) treatments as much as its empirical frequency distribution varied in their experiments.

Here we adapt our earlier methods to show that in VHBB's (1993) environment, strategic uncertainty generates robustness, optimistic subjects, and forward induction effects large enough to explain the pattern of equilibrium selection they observed. Our approach yields a unified explanation of VHBB's (1990, 1991, 1993) results, and makes it possible to assess the likely importance of those effects in other environments.

We assume throughout that players are rational in the standard sense that their decisions maximize expected payoffs, given their beliefs about each other's decisions. But our analysis is adaptive in that, rather than assuming equilibrium in the stage game or the repeated game that describes players' entire interaction, we allow players to have diverse beliefs about each other's decisions in the stage game and model the process by which their beliefs and decisions converge as they learn to predict each other's decisions. Thus, our analysis allows equilibrium to be reached by unsophisticated learning from actual observation rather than introspective, strategically sophisticated reasoning.

We impose the structure needed for analysis by using special features of VHBB's environment to give a simple econometric characterization of strategic uncertainty and the learning dynamics, which allows a more informative analysis of the effects of strategic uncertainty than now seems possible for games in general. Players' initial beliefs and responses to new observations each period are perturbed by independently and identically distributed (henceforth, "i.i.d.") random shocks. The shocks represent strategic uncertainty in terms of the differences in players' learning rules, and their variances represent the initial level of strategic uncertainty and how it varies over time as players learn to forecast each other's responses. In effect each player has his own theory of coordination, which gives his initial beliefs and his interpretations of new observations an unpredictable component.

The resulting model is a Markov process with time-varying transition probabilities, the dynamics of which are driven by strategic uncertainty. Like our earlier models, it encompasses the leading alternative characterizations of strategic behavior in such settings: traditional analyses in which players' beliefs and decisions are focused from the start on a particular equilibrium, evolutionary analyses of the "long-run equilibria" of ergodic dynamics with small amounts of noise, and history-dependent learning with lock-in on a particular equilibrium in the limit. These are distinguished mainly by different values of the variances that represent the level of strategic uncertainty and how it varies over time.

As in our earlier analyses, we take the variances and certain other aspects of behavior to be exogenous behavioral parameters. Our characterization provides a framework within which to close the model by estimating them, using VHBB's (1993) experimental data.[7] As before, the estimated parameters generally satisfy the restrictions suggested by theory, but differ significantly from the values needed to justify an equilibrium analysis or an analysis of the long-run equilibria of ergodic dynamics. Instead they indicate large initial levels of strategic uncertainty, declining to zero over time. The decline is rapid enough to make the

[7] The need to proceed this way will come as no surprise to anyone who accepts Thomas Schelling's (1960) premise that the analysis of coordination is inherently partly empirical. Strategic uncertainty is crucial to understanding the dynamics, but the differences in subjects' beliefs cannot be reliably explained by theory alone because subjects were indistinguishable and had nearly identical information.

dynamics converge with probability one to one of the pure-strategy subgame-perfect equilibria of the stage game, so that the model's implications for equilibrium selection can be summarized by the prior probability distribution of that equilibrium. The estimated model gives an adequate statistical summary of individual subjects' behavior, and implies probability distributions of the dynamics and limiting outcome that resemble their empirical frequency distributions in VHBB's experiment.

We study equilibrium selection in more detail by using special features of our learning model and the environment to obtain a closed-form solution for the histories of players' beliefs, bids, and efforts as functions of the behavioral parameters, the treatment variables, and the shocks that represent strategic uncertainty. The solution shows how the limiting outcome is built up period by period from the shocks, the effects of which persist indefinitely. This persistence makes the learning process resemble a random walk in the aggregate, but with declining variances and nonzero drift that depend on the parameters and treatment variables. As in our earlier analyses, large, persistent effects of strategic uncertainty make it necessary to analyze the entire learning process to understand equilibrium selection, and preclude explanations based on equilibrium refinements or ergodic dynamics, which assume away either strategic uncertainty or persistence.

The form of our solution indicates that unless the behavioral parameters vary sharply and unpredictably with changes in the environment—which our estimates for VHBB's (1990, 1991) experiments suggest is unlikely—the dynamics and limiting outcome will vary with the treatment variables in stable, predictable ways. We begin with a qualitative comparative dynamics analysis of the effects of changes in VHBB's (1993) treatment variables. Generalizing results from our earlier analyses, we make precise, in the probabilistic sense appropriate to the model, the common intuitions that coordination tends to be less efficient the closer the order statistic is to the minimum (our robustness effect) and less efficient in larger groups because it requires coherence among a larger number of independent decisions. We also establish a new result showing that coordination tends to be more efficient the more intense the competition for the right to play.

Our model explains the efficiency-enhancing effect of auctions as the result of a dynamic interplay between forward induction, optimistic subjects, and robustness effects. Our solution allows a quantitative comparative dynamics analysis that shows how the magnitudes of these effects are determined by the treatment variables and behavioral parameters, and makes it possible to estimate their importance in other environments. The mean coordination outcome can be expressed as the sum of four components per period, one of which is the forward induction effect, one of which combines the optimistic subjects effect with the robustness effect of our earlier analyses, and two of which are smaller effects, discussed below. These components can be approximated as known functions of the treatment variables, the behavioral parameters, and tabulated statistical parameters.

These approximations yield a simple characterization of the optimistic subjects and robustness effects. Together they have approximately the same magnitude in VHBB's environment (where the right to play a nine-person median game was auctioned in a group of 18) as the robustness effect in an 18-person coordination game without auctions in which payoffs and best responses are determined by the fifth highest (the median of the nine highest) of all 18 players' efforts. In this respect the auctions effectively transformed VHBB's median game, which without auctions would have a robustness effect that contributes zero drift to the dynamics, into a 75th-percentile game (0.75 = 13.5/18) with a robustness effect that contributes a large upward drift. Our estimates suggest that this drift is responsible for roughly half of the efficiency-enhancing effect of auctions in VHBB's environment, and that the other half is due to a strong forward induction effect.

Our quantitative analysis also shows that, unless the behavioral parameters vary sharply with changes in the environment, there is a positive efficiency-enhancing effect through

our the class of environments we study, but it is not always strong enough to assure fully efficient coordination. More generally, the mechanism suggests that the effect will extend to other laboratory or field environments that combine significant strategic uncertainty with the ingredients of our optimistic subjects and forward induction effects.

The rest of the paper is organized as follows. Section I describes the environments we study and introduces our learning model. Sections II and III present the theoretical and econometric analyses. Section IV is the conclusion.

I. The Model

In this section we describe the class of environments we study and introduce the learning model that is the basis for our theoretical and econometric analyses. We call the highest price at which a player plans to remain in the auction his *bid*, and we define *efficiency* with reference to players' payoffs in the coordination game, bearing in mind that the auction may transfer some or all of the surplus to the experimenters.

A. *The Environment*

In VHBB's (1993) experiment subjects played the nine-person median coordination game of VHBB's (1991) treatment Γ for 10 or 15 periods, with the right to play auctioned each period in a population of 18 subjects. In the median game nine players choose among seven efforts, {1, ... , 7}, with payoffs determined by their own efforts and the group median effort. Denoting players' efforts at time t by $\hat{x}_{1t}, \ldots, \hat{x}_{nt}$ and the median by M_t, player i's period-t payoff in dollars is $0.1M_t - 0.05(M_t - \hat{x}_{it})^2 + 0.6$. Because no player's effort can influence the median when all other players are choosing the same effort, and player i's payoff is higher, other things equal, when $\hat{x}_{it} = M_t$, any configuration of efforts with $\hat{x}_{it} = M_t$ for all i is an equilibrium, and these seven symmetric equilibria are the only pure-strategy equilibria. These equilibria are strict and Pareto ranked, with equilibria with higher M_t better for all players than those with lower M_t.

The auction was a multiple-unit ascending-bid English clock as in Kevin McCabe et al. (1990), in which subjects indicated their willingness to pay the current asking price by holding up bid cards, with subjects who dropped out not allowed to reenter the bidding. The asking price started five cents below the payoff of the least efficient equilibrium of the median game and increased by five-cent increments until 11 or fewer subjects remained, after which it increased by one-cent increments until nine or fewer subjects remained. (The payoffs of adjacent equilibria differed by ten cents.) The market-clearing price was determined as follows. If the lowest price at which nine or fewer subjects remained in the auction left exactly nine subjects, all nine were awarded the right to play at that price. If that price left fewer than nine subjects, they were all awarded the right to play, with the remainder of the nine slots filled randomly from those who dropped out at the last increase, and all nine subjects paying the price before the last increase.[8]

In VHBB's (1993) design, as in their treatment Γ, explicit communication was prohibited throughout; the median was publicly announced after each play; and with minor exceptions the structure was publicly announced at the start. The market-clearing price was publicly announced after each auction, before the winners played the median game.

We now describe a class of environments that generalizes VHBB's (1993) design by allowing any values of their treatment variables and any coordination game that shares the structural features noted above (including those used in VHBB's [1990, 1991] other treatments). We assume throughout that bids are continuously variable.[9] We introduce the model under the simplifying assumption that effort is also continuously variable. In Section I, subsection D, and more formally in Section III, subsection B, we explain how to adapt the

[8] Thus subjects never paid more than they indicated they were willing to, and they were never excluded involuntarily unless they had indicated approximate indifference.

[9] This was not literally true in the experiments, but the increments by which bids could vary were small enough (one cent near the market-clearing price, compared to a ten-cent payoff difference between equilibria) to make it a good approximation.

model to the discrete effort spaces of VHBB's experiment by viewing the continuously variable efforts as the latent variables in an ordered probit model of discrete effort choice.

There is a finite number, m, of indistinguishable players, who repeatedly play a stage game with symmetric player roles. The stage game consists of an n-person coordination game, with $n < m$, preceded by an auction in which all m players bid for the right to play.

The coordination game has one-dimensional strategy spaces, with strategies called efforts. Effort has a commonly understood scale, which makes it meaningful to say that a player chose the same effort in different periods, or that different players chose the same effort. Any symmetric effort combination is an equilibrium, and such combinations are the only pure-strategy equilibria. By symmetry, these equilibria are Pareto ranked unless players are indifferent between them. Each player's best responses are given by a summary statistic of all players' efforts when that statistic is unaffected by his effort. The summary statistic, y_t, is a function $f(x_{1t}, \dots, x_{nt})$, where x_{it} is player i's continuously variable effort at time t and players $1, \dots, n$ play the game. We assume that $f(\cdot)$ is continuous and, for any $x_{1t}, \dots, x_{nt}$ and constants a and $b \geq 0$, $f(a + bx_{1t}, \dots, a + bx_{nt}) \equiv a + bf(x_{1t}, \dots, x_{nt})$. These assumptions are satisfied when $f(\cdot)$ is an order statistic or a convex combination of order statistics such as the arithmetic mean. To see what they entail, note that the symmetry of roles implies that $f(\cdot)$ is a symmetric function of the x_{it}, so that its value is determined by the order statistics of their empirical distribution. Our assumptions rule out most nonlinear functions of these order statistics, which is restrictive but probably not unrepresentative of symmetric games. We abuse terminology by calling $f(\cdot)$ an order statistic.

The auction is a multiple-unit ascending-bid English clock as described above, with player i's bid and the market-clearing price at time t denoted p_{it} and q_t, respectively. We depart from the auction VHBB used only in assuming that p_{it} and q_t are continuously variable. With continuously variable bids the effects of VHBB's tie-breaking procedure are negligible, and q_t can be taken to be the $(n+1)$st largest of the p_{it}, an order statistic of their empirical distribution. For any given m and n, q_t can be written as a function,

$$(1) \qquad q_t \equiv g(p_{1t}, \dots, p_{mt}),$$

where for any $p_{1t}, \dots, p_{mt}$ and any constants a and $b \geq 0$,

$$(2) \qquad g(a + bp_{1t}, \dots, a + bp_{mt}) \equiv a + bg(p_{1t}, \dots, p_{mt}).$$

Because the form of $g(\cdot)$ is completely determined by m and n, we describe changes in $g(\cdot)$ below by describing the associated changes in m and n, without separate reference to $g(\cdot)$.

The outcome of the stage game can now be described more precisely. We write

$$(3) \qquad y_t \equiv h(x_{1t}, \dots, x_{mt}; p_{1t}, \dots, p_{mt}),$$

where the x_{it} now include planned effort choices for all m players, and $h(x_{1t}, \dots, x_{mt}; p_{1t}, \dots, p_{mt})$ equals $f(\cdot)$ evaluated at the x_{it} associated with the n highest p_{it} in the population—noting that the order of these x_{it} is immaterial and ignoring ties in the p_{it}, which will have probability zero under our distributional assumptions. It is immediate from the definition that for any $p_{1t}, \dots, p_{mt}$ and any constants a and $b \geq 0$, $h(\cdot)$ is continuous in $x_{1t}, \dots, x_{mt}$ and inherits the scaling properties of $f(\cdot)$, in that

$$(4) \qquad h(a + bx_{1t}, \dots, a + bx_{mt}; p_{1t}, \dots, p_{mt}) \equiv a + bh(x_{1t}, \dots, x_{mt}; p_{1t}, \dots, p_{mt}).$$

As only the order of the p_{it} matters, it is also clear that for any constants a and $b > 0$,

$$(5) \qquad h(x_{1t}, \dots, x_{mt}; a + bp_{1t}, \dots, a + bp_{mt}) \equiv h(x_{1t}, \dots, x_{mt}; p_{1t}, \dots, p_{mt}).$$

We assume that the structure of the environment is publicly known.

B. *Equilibrium Refinements in the Stage Game*

As noted above, in VHBB's (1993) experiment eight out of eight subject groups bid the market-clearing price of the right to play to a level recoverable only in the efficient equilibrium, and their efforts converged to that equilibrium. As VHBB's (1993) Tables V and VI make clear, subjects' bids and efforts were generally consistent with the intuition for forward induction, in that individual subjects seldom bid more than their efforts made it possible to recover in the coordination game, and never did so after the first few periods.

Although we shall argue below that the limiting outcomes in the experiment cannot be understood without analyzing history-dependent learning dynamics, it is instructive to consider the implications of equilibrium refinements in the stage game. Because the structure was publicly announced, we assume complete information. The most relevant refinements are subgame perfectness and forward induction, and the essential points can be seen by focusing on symmetric pure-strategy equilibria. Without loss of generality we rescale the payoffs in the coordination game, which are increasing in players' common efforts, so that in each of these equilibria players' payoffs equal their efforts.

When all players use the same bidding strategy and $n < m$, no player can alter the market-clearing price by changing his bid. It is easy to check that any strategy combination with $p_t = x_{it} = q_t = y_t$ for all i, in which each player bids his payoff in a symmetric pure-strategy equilibrium of the coordination game and then plays that equilibrium's effort unless he is the only player to bid that amount, is a subgame-perfect equilibrium of the stage game.[10]

In applying forward induction we follow Ben-Porath and Dekel (1992) in identifying it with the ability to survive iterated deletion of weakly dominated strategies. VHBB's stage game resembles Ben-Porath and Dekel's Figure 2.3b example, in which two players publicly and simultaneously decide whether to burn a given quantity of money before playing a coordination game like VHBB's.[11] This game has a symmetric subgame-perfect equilibrium in which players burn money and then play the inferior equilibrium in the coordination game, each anticipating that equilibrium unless he is the only player who burns money. In this equilibrium each player's strategy is a unique best response given his beliefs, so it survives iterated deletion of dominated strategies. As Ben-Porath and Dekel note, such equilibria normally exist whenever players move simultaneously in the communication phase. In VHBB's stage game such an equilibrium can be constructed to support any of the seven symmetric pure-strategy equilibria in their coordination game, with full surplus extraction in the auction. The requirements for such equilibria are also consistent with subgame perfectness. Thus, subgame perfectness and forward induction in the stage game, together or separately, are consistent with the limiting outcomes in VHBB's experiment but too unrestrictive to help in explaining them. We use "forward induction" loosely below to refer to the intuition rather than a formal definition.

C. *The Learning Model*

For our purposes, it is essential to capture the idea that, even if players form their beliefs and choose their bids and efforts sensibly, they may differ in unpredictable ways. It is also essential to describe the dynamics of beliefs,

[10] No player can gain by changing his bid because the right to play yields zero surplus. Other symmetric pure-strategy subgame-perfect equilibria differ from these only inessentially, off the equilibrium path. There are no symmetric equilibria in which the right to play yields positive surplus because a player could gain by outbidding his partners, assuring himself the right to play without raising the market-clearing price. Kim (1996 Lemma 1) gives a full characterization of the pure-strategy subgame-perfect equilibria in this game, including asymmetric ones.

[11] VHBB's stage game differs from Ben-Porath and Dekel's example in that the latter's players must play the game whether or not they burn money, and must bear the cost of any money they burn no matter what their partners do; but these differences are inessential here. Note that unilateral deviations from symmetric bid combinations are always observable, although deviations from asymmetric combinations might not be.

bids, and efforts realistically in terms of observable variables, in a way that permits estimation of the behavioral parameters that cannot reliably be determined by theory (such as the variances that represent the level of strategic uncertainty) and allows an informative theoretical analysis.

In the environments we study, players' bids and efforts evolve as follows. First all m players choose their initial bids, the resulting value of q_t is publicly announced, and the n winners choose their efforts in the coordination game.[12] All m players then observe the resulting value of y_t and choose new bids, and the process continues. Because the structure is public knowledge, players face uncertainty only about each other's bids and efforts. The effects of others' bids and efforts on a player's payoffs are filtered through q_t and y_t.

The large subject populations in VHBB's (1993) experiment make it a plausible working hypothesis that players treat their individual influences on q_t and y_t as negligible. This implies that players' optimal bids and efforts each period are determined by their current payoff implications, and thus by their beliefs about the current y_t, because those beliefs determine their expected payoffs and optimal efforts, and their bids make winning the right to play contingent on q_t. We imagine that players begin with prior beliefs about the process that generates q_t and y_t, use standard statistical procedures to revise their beliefs in response to new observations, and choose the bids and efforts that are optimal, given their beliefs. Players whose priors differ may then have different beliefs even if they have always observed the same history and used the same procedures to interpret it.

We adopt a simple econometric characterization of the dynamics of beliefs, bids, and efforts in the style of the adaptive control literature.[13] The key insight of the control literature is that describing how beliefs respond to new information does not require representing them as probability distributions or their moments. It is enough to model the dynamics of the optimal decisions they imply, which is the only aspect of beliefs that directly affects the outcome. We represent players' beliefs by their optimal efforts, which when continuous variable preserve enough information to realistically describe their bids as well as their efforts.[14] There is no need to assume that optimal efforts are determined in any simple way by the moments of the distributions that describe players' beliefs.

We describe the conditional means of players' responses to new observations by simple linear adjustment rules, in which their beliefs adjust part of the way toward the value suggested by the latest observation of y_t in a way that generalizes the fictitious-play and best-response rules to allow different values of parameters that represent the initial levels, trends, and inertia of beliefs. We describe the differences in players' beliefs by perturbing these mean adjustments by idiosyncratic random shocks each period, which represent strategic uncertainty described in terms of the differences in players' learning rules. We assume that these shocks are i.i.d. across players with zero means and given variances. Thus any correlation in players' beliefs that emerges is the result of responses to common observations of y_t and q_t rather than an artifact of our distributional assumptions.

We distinguish between players' beliefs when they bid and when the winners in the auction play the coordination game, because the latter beliefs may reflect inferences from q_t. Player i's beliefs in period t before and after he observes q_t are denoted $\bar{x}_{it}$ and x_{it}, respectively. They are interpreted as his estimates of his optimal effort given his information at those times, with $\bar{\mathbf{x}}_t$ and $\mathbf{x}_t$ denoting the associated population m-vectors.

[12] We assume that bids are not publicly observable, even though VHBB's subjects were in booths that may have allowed them, with some effort, to observe when other subjects dropped out. At our request Weber (1994) ran two trials that replicated VHBB's design but ensured that bids were unobservable, with very similar results.

[13] See M. B. Nevel'son and R. Z. Has'minskii (19[illegible]) or Michael Woodford (1990 Sec. 2).

[14] Our model with discrete efforts, by contrast, describes beliefs and efforts separately.

Players' initial beliefs are given by

$$(6) \qquad \bar{x}_{i0} = \alpha_0 + \zeta_{i0},$$

where α_0 is their common mean and the ζ_{i0} are idiosyncratic shocks.

Recalling that payoffs in the coordination game are measured so that players' payoffs in symmetric equilibria equal their efforts, players' bids in period t are given by

$$(7) \qquad p_{it} = \beta_t + \bar{x}_{it} + \eta_{it}, \qquad t = 0, \ldots,$$

where the β_t are constants and the η_{it} are idiosyncratic shocks that represent the differences in players' bidding strategies. This specification can be justified as follows. Players' values of the right to play may be affiliated, because with strategic uncertainty higher bids signal higher beliefs and efforts, which raise the *ex ante* distributions of all players' payoffs.[15] But because players ignore their future influences on y_t, and their individual influences on q_t are negligible unless they do not win the right to play, standard arguments show that it is approximately optimal for a player to stay in the auction until his value given the current level of q_t falls below q_t [although q_t influences the outcome in the game, via (8) below, this criterion uniquely defines his bid when $\delta_t < 1$].[16] Finally, because VHBB's median game imposes a quadratic payoff penalty for efforts away from the median, a risk-neutral player's optimal effort equals the mean of his distribution of y_t by certainty-equivalence. His expected payoff is then equal to that effort minus the variance of his distribution of y_t, which should be approximately proportional to the variance of θ_{it} in (8) below. This yields (7), with β_t, the mean of players' corrections for the cost of strategic uncertainty, negative and approaching zero over time if Var $\theta_{it} \to 0$.

After the auction players observe q_t, as determined by (1), and update their beliefs to

$$(8) \qquad x_{it} = \gamma_t + \delta_t q_t + (1 - \delta_t)\bar{x}_{it} + \theta_{it}, \qquad t = 0, \ldots,$$

where the γ_t and δ_t are constants and the θ_{it} are shocks that represent the differences in updating rules. γ_t and δ_t, the level and slope parameters of these linear adjustments, together reflect forward induction inferences from the observed value of q_t.

The winners then determine their efforts according to the x_{it}, and all m players observe the value of y_t determined by (3) and update their beliefs again, to

$$(9) \qquad \bar{x}_{it+1} = \alpha_{t+1} + \varepsilon_{t+1} y_t + (1 - \varepsilon_{t+1}) x_{it} + \zeta_{it+1}, \qquad t = 0, \ldots,$$

where the α_t and ε_t are constants and the ζ_{it} are shocks. The level and slope parameters α_{t+1} and ε_{t+1} represent any trends in beliefs and the precision of beliefs, as reflected by how much they respond to new observations of y_t. The $\bar{x}_{it+1}$ then determine players' bids next period via (7), and the process continues.

Although (8) and (9) suggest partial adjustment to the efforts suggested by the latest observation of q_t or y_t, they are best thought of as representing full adjustment to players' current estimates of their optimal efforts, which respond less than fully to new observations because they are only part of players' information about the process. As explained in Crawford (1995 pp. 112–13), (9) generalizes the familiar fictitious-play and best-

[15] Paul Milgrom and Robert Weber (1982) define affiliation. Even though winners play the same game, their values are not common because their beliefs and efforts differ.

[16] VHBB's auction is a multiple-unit version of the "Japanese" auction studied by Milgrom and Weber (1982). They show that when players have affiliated values and can observe each other's bids, it is optimal for a player to stay in until the current price makes him indifferent between winning or losing, taking into account the winner's curse. However, in VHBB's auctions winning reveals nothing to a player about other winners' beliefs; and anything winners learn about losers' beliefs is irrelevant to the current value of the game unless the idiosyncratic parts of players' beliefs are correlated. This suggests that winners' curse effects are unimportant here, and that it makes little difference whether players can observe losers' bids.

response learning rules to allow a much wider range of priors about the structure of the y_t process.[17] Equations (8) and (9) are not fully "rational" in the game-theoretic sense, because players' priors are not required to be correct and linearity may be inconsistent with the rules that are optimal given their priors. However, in the control literature such rules have been shown to provide a robust estimation method for agents who understand the forecasting problems they face, but are unwilling to make the specific assumptions about the structure of the process or unable to store and process the large amounts of information a Bayesian approach requires.

By construction, $E\zeta_{it} = E\eta_{it} = E\theta_{it} = 0$ for all i and t, where E is the expectation operator. We assume that ζ_{it}, η_{it}, and θ_{it} are independent of each other and serially independent for each i, which amounts to assuming that $\bar{x}_{it}$ and x_{it} fully capture the future effects of past idiosyncratic influences on player i's beliefs. We also assume that ζ_{it}, η_{it}, and θ_{it} are i.i.d. across i, with common variances denoted $\sigma^2_{\zeta t}$, $\sigma^2_{\eta t}$, and $\sigma^2_{\theta t}$.

We assume throughout that $0 < \delta_t, \varepsilon_t \leq 1$. With no further restrictions on the behavioral parameters, the model is formally consistent with any history of the y_t for any m, n, and $f(\cdot)$, with α_t varying over time as necessary and β_t, γ_t, $\sigma^2_{\zeta t}$, $\sigma^2_{\eta t}$, and $\sigma^2_{\theta t}$ set equal to zero, so that for all i and t, $p_{it} = \bar{x}_{it} = x_{it} = q_t = y_t = \Sigma^t_{s=0}\,\alpha_s$. Such solutions, in which players jump from one stage-game equilibrium to the next following some commonly understood pattern, are not ruled out by standard arguments because they correspond to equilibria of the repeated game. Yet they are empirically bizarre, because the ad hoc time variation in the behavioral parameters they require violates the hypothesis that, in a given environment, the distribution of past behavior is indicative of current behavior, which most of us take for granted in learning to predict others' behavior in repeated interactions.

We address this issue below by imposing intertemporal restrictions that rule out such ad hoc variations in the model's behavioral parameters. These restrictions allow for the possibility that α_t, β_t, γ_t, $\sigma^2_{\zeta t}$, $\sigma^2_{\eta t}$, $\sigma^2_{\theta t} \to 0$ as $t \to \infty$, so that players learn to predict y_t if it converges and choose the x_{it} that are optimal given their predictions, and the long-run steady states of the dynamics coincide with the pure-strategy subgame-perfect equilibria of the stage game [shown in Section I, subsection B, to satisfy $p_{it} = x_{it}$ $(= \bar{x}_{it}) = q_t = y_t$ for all i]. We find below that they permit an adequate econometric description of individual subjects' behavior, which yields a useful explanation of the dynamics of their interactions.

D. *Discrete Efforts*

To estimate the model and to use it to predict experimental results, we must allow for the discrete efforts of VHBB's design. We do this by letting x_{it} and $\bar{x}_{it}$ continue to represent beliefs and remain continuously variable, but with x_{it} determining effort as the latent variable in an ordered probit, described in Section III, subsection B. Beliefs, bids, and efforts still follow (1), (3), and (6)–(9), but with efforts and the order statistic determined by rounding each x_{it} to the nearest feasible integer and evaluating $h(\cdot)$ at the rounded values.

II. Analysis

We begin by analyzing the model with continuously variable efforts. In Section II, subsection F, we explain how to extend the analysis to the discrete efforts of VHBB's experiment.

A. *Preliminaries*

Once the distributions of ζ_{it}, η_{it}, and θ_{it} are specified, (1), (3), and (6)–(9) define a Markov process with state vector $\bar{x}_t$, in which beliefs, bids, and efforts are i.i.d. across players *ex ante*. Equation (6) provides initial conditions and substituting (1

[17] Unlike learning rules in which players respond only to realized payoffs, such as those in Alvin Roth and Ido Erev (1995), (8) and (9) reflect the best-response structure. As explained in Crawford (1997 Sec. 6.3), VHBB's subjects seemed to understand the structure, and such rules allow a better description of their behavior.

(3), and (7)–(8) into (9) and using (2) and (4)–(5) gives $\bar{x}_{it+1}$ as a function of the $\bar{x}_{it}$, parameters, and shocks:

$$
\begin{aligned}
(11)\quad \bar{x}_{it+1} = {} & \alpha_{t+1} + \beta_t\delta_t + \gamma_t \\
& + \delta_t g(\bar{x}_{1t} + \eta_{1t}, \dots, \bar{x}_{mt} + \eta_{mt}) \\
& + \varepsilon_{t+1} h[(1-\delta_t)\bar{x}_{1t} + \theta_{1t}, \dots, \\
& (1-\delta_t)\bar{x}_{mt} + \theta_{mt}; \bar{x}_{1t} + \eta_{1t}, \dots, \bar{x}_{mt} + \eta_{mt}] \\
& + (1-\varepsilon_{t+1})[(1-\delta_t)\bar{x}_{it} + \theta_{it}] \\
& + \zeta_{it+1}, \qquad t = 0, \dots .
\end{aligned}
$$

The model's recursive structure and the independence of players' deviations from the average adjustment rules capture the requirement that players form their beliefs and choose their strategies independently that is the essence of the coordination problem.

The model's dynamics are driven by strategic uncertainty, as represented by the $\sigma^2_{\zeta t}$, $\sigma^2_{\eta t}$, and $\sigma^2_{\theta t}$. Different assumptions about how $\sigma^2_{\zeta t}$, $\sigma^2_{\eta t}$, and $\sigma^2_{\theta t}$ vary over time have different implications for the dynamics, which go a long way toward identifying the stochastic structure. The following result is helpful in understanding this relationship.

PROPOSITION 0: *Suppose that $\alpha_t = 0$ for all $t = 1, \dots$ and $\beta_t = \gamma_t = 0$ for all $t = 0, \dots$, so that there are no exogenous trends in beliefs, and that $\sigma^2_{\eta t} = \sigma^2_{\theta t} = \sigma^2_{\zeta t} = 0$ for all $t = T, \dots$, so that the rules that describe players' responses to new observations coincide from period T onward for some $T \geq 0$. Then there exist players, j and k, such that $g(\bar{x}_{1t}, \dots, \bar{x}_{mt}) = \bar{x}_{jt}$ and $h(\bar{x}_{1t}, \dots, \bar{x}_{mt}; \bar{x}_{1t}, \dots, \bar{x}_{mt}) = \bar{x}_{kt}$ for all $t = T, \dots$; and $\bar{x}_{it}$ and x_{it} converge with probability one to a common value, weakly between $\bar{x}_{jT}$ and $\bar{x}_{kT}$, that is completely determined by $\bar{x}_{jT}$, $\bar{x}_{kT}$, and the δ_t and ε_t for $t = T, \dots$. If $T = 0$, so that players' initial beliefs coincide as well, then $\bar{x}_{it} = x_{it} = \alpha_0$ for all i and all $t = 0, \dots$.*

The proof is in the Appendix. Proposition 0 shows that if there are no differences in players' responses to new information from period T on, the dynamics must converge to an equilibrium dictated by, and bounded between, the $\bar{x}_{jT}$ and $\bar{x}_{kT}$ that determine $g(\cdot)$ and $h(\cdot)$ in period T. In VHBB's environment $\bar{x}_{jT}$ and $\bar{x}_{kT}$ are the tenth and fifth largest $\bar{x}_{iT}$ (because $n = 9$, $g(\cdot)$ is the $(n + 1)$th highest bid, and $h(\cdot)$ is the median of the nine largest $\bar{x}_{iT}$). Thus, unless differences in responses persist long enough to drive $\bar{x}_{jT}$ and $\bar{x}_{kT}$ to efficient levels, the model cannot explain the convergence to the efficient equilibrium VHBB observed.[18]

If $\sigma^2_{\eta t}$, $\sigma^2_{\theta t}$, and $\sigma^2_{\zeta t}$ converge to positive constants, instead of remaining at 0 for all $t = T, \dots$, an analysis of long-run equilibria seems possible, along the lines of the analysis of evolutionary dynamics in Kim (1996). This would probably reproduce Kim's conclusion that there is efficient coordination in the long run, as in VHBB's experiment. However, the closely related experiments of Weber (1994) and Cachon and Camerer (1996) suggest that efficiency is not inevitable for all values of m, n, and $f(\cdot)$ (although no amount of data from experiments with finite horizons can ever disprove a long-run result). More importantly, we find below that when $\sigma^2_{\eta t}$, $\sigma^2_{\theta t}$, and $\sigma^2_{\zeta t}$ decline gradually to 0, as one would expect as players learn to forecast q_t and y_t from their common observations, the model yields history-dependent dynamics that closely resemble VHBB's results. And in this case the model also provides useful information about the mechanism behind the efficiency-enhancing effect of auctions and the effects of changes in the environment—information that would very likely be lost in an analysis of long-run equilibria.

B. *Closed-Form Solution*

We now show how to analyze the dynamics, whether or not the variances converge to zero, in a way that yields a deeper understanding of

[18] This not the whole story because in our empirical analysis α_t, β_t, and γ_t sometimes differ from 0 and, as explained below, these parameters also affect efficiency. We set them equal to 0 here, except for α_0, to focus on a more subtle aspect of the dynamics.

how outcomes are determined and makes it possible to estimate the effects of changes in the environment.

The key to much of our analysis is the fact that the scaling properties of the order statistics $g(\cdot)$ and $h(\cdot)$ and the form of players' learning rules allow us, despite the model's nonlinearity, to obtain a closed-form solution for the entire history of players' beliefs, bids, and efforts as a function of the behavioral parameters and the shocks that represent strategic uncertainty. In what follows, sums with no terms (like $\Sigma_{s=0}^{t-1} \varepsilon_{s+1} h_s$ for $t = 0$) are understood to equal 0, and products with no terms are understood to equal 1.

PROPOSITION 1: *The unique solution of the learning dynamics is given, for all i and t, by*

$$(12) \quad x_{it} = \sum_{s=0}^{t} [\alpha_s + \beta_s \delta_s + \gamma_s] + \sum_{s=0}^{t} \delta_s g_s + \sum_{s=0}^{t-1} \varepsilon_{s+1} h_s + z_{it}$$

and

$$(13) \quad y_t = \sum_{s=0}^{t} [\alpha_s + \beta_s \delta_s + \gamma_s] + \sum_{s=0}^{t} \delta_s g_s + \sum_{s=0}^{t-1} \varepsilon_{s+1} h_s + h_t,$$

where

$$(14) \quad g_t \equiv g[\bar{z}_{1t} + \eta_{1t}, \ldots, \bar{z}_{mt} + \eta_{mt}],$$

$$h_t \equiv h(z_{1t}, \ldots, z_{mt}; \bar{z}_{1t} + \eta_{1t}, \ldots, \bar{z}_{mt} + \eta_{mt}),$$

$$(15) \quad z_{it} \equiv \sum_{s=0}^{t} [(1 - \delta_s)\zeta_{is} + \theta_{is}] \times \left[\prod_{j=1}^{t-s} (1 - \delta_{t-j+1})(1 - \varepsilon_{t-j+1}) \right],$$

and $\bar{z}_{it} \equiv (1 - \varepsilon_t) z_{it-1} + \zeta_{it}$.

The proof is in the Appendix. Proposition 1 expresses the outcome of the learning process as the cumulative effect of trend terms α_t, $\beta_t\delta_t$, and γ_t and shock terms $\delta_t g_t$ and $\varepsilon_{t+1} h_t$, the effects of which persist indefinitely.[19] This persistence makes the process resemble a random walk in the aggregate, but with nonzero drift and declining variances determined by the parameters and the treatment variables. As explained in Section I, subsection C, α_0 and the α_t represent the average level of players' initial beliefs and trends in their beliefs, the β_t correct their bidding rules for the expected cost of strategic uncertainty, and the γ_t reflect their average forward induction inferences on observing q_t. The $\delta_t g_t$ reflect the effect of dispersion in players' bids on their beliefs via the market-clearing price p_t. The $\varepsilon_{t+1} h_t$ reflect the robustness effect of the dispersion of the winners' efforts via y_t, characterized in Broseta (1993a Proposition 1) and Crawford (1995 Proposition 1), modified here by the optimistic subjects effect. Proposition 1 is valid no matter how the shocks are generated, but much of its usefulness stems from the fact that the shock terms are known functions of the z_{it}, $\bar{z}_{it}$, and η_{it}, which are *ex ante* i.i.d. across i for any t. This and the scaling properties of $g(\cdot)$ and $h(\cdot)$ are crucial to the comparative dynamics analysis below.

C. *Convergence*

In the next proposition, for technical reasons we bound players' efforts. This is done by increasing x_{it} to its lower bound, denoted $\underline{x}$, or reducing it to its upper bound, denoted $\bar{x}$, when it would otherwise fall outside the interval $[\underline{x}, \bar{x}]$.

PROPOSITION 2: *Assume that the x_{it} are bounded, so that they remain in the interval $[\underline{x}, \bar{x}]$. If $\Sigma_{s=0}^{\infty} \alpha_s$, $\Sigma_{s=0}^{\infty} \beta_s$, $\Sigma_{s=0}^{\infty} \gamma_s$, $\Sigma_{s=0}^{\infty}$ $\sigma_{\theta s}^2$, $\Sigma_{s=0}^{\infty} \sigma_{\eta s}^2$, and $\Sigma_{s=0}^{\infty} \sigma_{\zeta s}^2$ are finite, and $0 < \delta_t, \varepsilon_t \leq 1$ for all t, with $(1 - \varepsilon_t)(1 - \delta_t)$ bounded below 1 for sufficiently large t, then*

[19] The remaining terms, z_{it} in (12) and h_t in (13), are subsumed in the shock terms after the period in which they first appear; and the remaining endogenous variables, p_{it}, and q_t, are easily computed from x_{it-1} and y_{t-1} via (1), (7), and (9).

(iii) the y_t and the $\bar{x}_{it}$ and x_{it} converge, with probability one, to a common, finite limit.

The proof is in the Appendix. Proposition 2 gives conditions under which the learning dynamics converge to the bids and efforts associated with one of the symmetric pure-strategy subgame-perfect equilibria of the stage game. The model's implications for equilibrium selection are best summarized by the prior probability distribution of that equilibrium, which by Proposition 1 is typically nondegenerate due to the persistent effects of strategic uncertainty. Although Proposition 2 is quite helpful in understanding the model, its conditions are for convergence with probability one, and are only sufficient; convergence can easily occur when they are violated. The variance conditions are as one would expect from the strong law of large numbers, and probably cannot be weakened significantly if one insists on convergence with probability one. However, the estimated model reported in Section III, subsection E, converges strongly to the upper bound $\bar{x}$ in VHBB's environment, even though our parameter estimates violate Proposition 2's conditions on the β_s and γ_s. This suggests that our conditions on the α_s, β_s, and γ_s could be weakened, but the additional insight this would yield seems unlikely to justify the effort.

D. *Qualitative Comparative Dynamics*

Proposition 1 shows that unless the behavioral parameters vary sharply with changes in the treatment variables, the dynamics and limiting outcome respond to such changes in stable, predictable ways. We now conduct a qualitative comparative dynamics analysis of changes in m, n, and $f(\cdot)$, holding the behavioral parameters constant.

We define the jth order statistic of an empirical distribution $(\xi_1, \ldots, \xi_n)$ as the jth smallest of the ξ_i; thus, the minimum is the first order statistic and the median (for odd n) is the $((n+1)/2)$th. If $f(\cdot)$ is an order statistic, the index j identifies which one; and if $f(\cdot)$ is a convex combination of order statistics, j indexes the order statistics from which it is computed. An "increase in j" refers to any shift in the weights used to compute $f(\cdot)$ that increases j in the sense of first-order stochastic dominance, and a random variable is said to "stochastically increase" if its probability distribution shifts upward in this sense.

PROPOSITION 3: *Holding m and n and the behavioral parameters constant, increasing j stochastically increases y_t and x_{it} for all i and t.*

PROOF:

The proof is immediate from Proposition 1, noting that for any history of shocks, increasing j increases h_t for all t and leaves the g_t unaffected.

PROPOSITION 4: *Holding m and j (or the weights on alternative values of j) and the behavioral parameters constant, increasing n stochastically decreases y_t and x_{it} for all i and t.*

PROOF:

The proof is immediate from Proposition 1, noting that for any history of shocks, increasing n (with i running from 1 to the larger value of n) decreases both g_t and h_t for all t.

PROPOSITION 5: *Holding n and j (or the weights on alternative values of j) and the behavioral parameters constant, increasing m stochastically increases y_t and x_{it} for all i and t.*

PROOF:

The proof follows from Proposition 1. It is clear that for any history of shocks, increasing m increases g_t. It is not true that increasing m increases h_t for any history, because such an increase may cause turnover among those who win the right to play. However, it follows from (8) and the fact that the θ_{it} are i.i.d. with mean zero that any turnover replaces players with players whose effort distributions first order stochastically dominate them (whether or not the x_{it} are bounded), which stochastically increases h_t.

Propositions 3 and 4 extend Broseta (1993a Lemma 3.1) and Crawford (1995 Propositions 3–4), making precise, in the probabilistic sense appropriate to the model, the intuitions that coordination is less efficient when more

efficient outcomes are robust only to deviations by smaller sets of players, and less efficient in larger groups because it requires coherence among a larger number of independent decisions. Proposition 5 addresses a new issue, showing that increased competition for the right to play favors efficiency because it tends to yield higher market-clearing prices and intensifies the optimistic subjects effect.

E. *Quantitative Comparative Dynamics*

We now conduct a quantitative comparative dynamics analysis, showing how the mean coordination outcome is affected by changes in the treatment variables and the behavioral parameters. The analysis requires some additional notation. Let σ_{pt}^2 and σ_{zt}^2, respectively, denote the common *ex ante* variances of the p_{it} (or, equivalently, the $\bar{z}_{it} + \eta_{it}$) and the z_{it}. Because the ζ_{it} and θ_{it} are serially independent, (15) and (7) imply that

$$(16)\quad \sigma_{zt}^2 \equiv \sum_{s=0}^{t} \left[(1-\delta_s)^2 \sigma_{\zeta s}^2 + \sigma_{\theta s}^2\right] \times \left[\prod_{j=1}^{t-s} (1-\delta_{t-j+1})(1-\varepsilon_{t-j+1})\right]^2$$

and

$$(17)\quad \sigma_{pt}^2 \equiv (1-\varepsilon_t)^2 \sigma_{z,t-1}^2 + \sigma_{\zeta t}^2 + \sigma_{\eta t}^2.$$

Finally, let $\mu_t \equiv Eg((\bar{z}_{1t} + \eta_{1t})/\sigma_{pt}, \dots, (\bar{z}_{mt} + \eta_{mt})/\sigma_{pt})$ and $\nu_t \equiv Eh(z_{1t}/\sigma_{zt}, \dots, z_{nt}/\sigma_{zt}; (\bar{z}_{1t} + \eta_{1t})/\sigma_{pt}, \dots, (\bar{z}_{mt} + \eta_{mt})/\sigma_{pt})$. Because the random variables $(\bar{z}_{it} + \eta_{it})/\sigma_{pt}$ and z_{it}/σ_{zt} are standardized, with mean 0 and variance 1, μ_t and ν_t are purely statistical parameters, completely determined by the joint distribution of these random variables and the treatment variables m, n, and $f(\cdot)$; they are subscripted because that distribution is time-dependent, and their dependence on m, n, and $f(\cdot)$ is suppressed for clarity. μ_t and ν_t respond to changes in m, n, and $f(\cdot)$ in the expected directions, as can be inferred from the responses of g_t and h_t identified in the proofs of Propositions 3–5.

PROPOSITION 6: *The ex ante means of the y_t and the x_{it} are given, for all i and t, by*

$$(18)\quad Ex_{it} = \sum_{s=0}^{t} [\alpha_s + \beta_s\delta_s + \gamma_s] + \sum_{s=0}^{t} \delta_s\sigma_{ps}\mu_s + \sum_{s=0}^{t-1} \varepsilon_{s+1}\sigma_{zs}\nu_s$$

and

$$(19)\quad Ey_t = \sum_{s=0}^{t} [\alpha_s + \beta_s\delta_s + \gamma_s] + \sum_{s=0}^{t} \delta_s\sigma_{ps}\mu_s + \sum_{s=0}^{t-1} \varepsilon_{s+1}\sigma_{zs}\nu_s + \sigma_{zt}\nu_t.$$

PROOF:

Using (2), (4), (5), and (14) yields

$$(20)\quad Eg_s \equiv E[\sigma_{ps}g((\bar{z}_{1s} + \eta_{1s})/\sigma_{ps}, \dots, (\bar{z}_{ms} + \eta_{ms})/\sigma_{ps})] \equiv \sigma_{ps}\mu_s$$

and

$$(21)\quad Eh_s \equiv E[\sigma_{zs}Eh(z_{1s}/\sigma_{zs}, \dots, z_{ns}/\sigma_{zs}; (\bar{z}_{1s} + \eta_{1s})/\sigma_{ps}, \dots, (\bar{z}_{ms} + \eta_{ms})/\sigma_{ps})] \equiv \sigma_{zs}\nu_s.$$

Taking expectations in (12) and (13) then yields (18) and (19).

Proposition 6 expresses the mean coordination outcome as a simple function of the behavioral parameters; the σ_{pt} and σ_{zt}, which can be computed from the behavioral parameters using (16) and (17); and the statistical parameters μ_t and ν_t, which reflect the influence of m, n, and $f(\cdot)$ on the dynamics. To apply this result we must estimate the behavioral parameters and evaluate μ_t and ν_t. The estimation is carried out in Section III. μ_t and ν_t are difficult to evaluate because their dependence on the distribution of the $\bar{z}_{it} + \eta_{it}$ and z_{it} is comp

but it is possible to approximate them, as we now describe.

The approximations are based on further assumptions. The first addresses a difficulty in evaluating Eh_t that arises because unless $\sigma^2_{\eta t} = \sigma^2_{\theta t} = 0$, so that there is no dispersion in players' bidding rules and their inferences on observing q_t, the highest bidders need not be those who would choose the highest efforts in the coordination game. We assume that

$$(22)\quad Eh_t \equiv Eh(z_{1t}, \dots, z_{mt};$$

$$\bar{z}_{1t} + \eta_{1t}, \dots, \bar{z}_{mt} + \eta_{mt})$$

$$\approx Eh(z_{1t}, \dots, z_{mt}; z_{1t}, \dots, z_{mt}),$$

so that the increases and decreases in $h(\cdot)$ caused by such "crossovers" approximately cancel out on average. This should not be an important source of bias, because

$$(23)\quad z_{it} = (1 - \delta_t)\bar{z}_{it} + \theta_{it}$$

$$= (1 - \delta_t)(\bar{z}_{it} + \eta_{it})$$

$$- (1 - \delta_t)\eta_{it} + \theta_{it},$$

so that there are no systematic differences between the orderings of the z_{it} and $\bar{z}_{it} + \eta_{it}$.

This approximation yields a simple characterization of the optimistic subjects effect, in which the order statistic is adjusted to reflect the auction's tendency to select players who choose higher efforts. In VHBB's environment, where the right to play a nine-person median game was auctioned in a group of 18, Eh_t is approximated by the robustness effect in an 18-person game without auctions in which the order statistic is the fifth highest effort (the median of the nine highest efforts). Thus, for the purpose of evaluating Eh_t the auctions effectively transformed VHBB's median game, which without auctions would have a robustness effect that contributes zero drift to the dynamics, into a 75th-percentile game ($0.75 = 13.5/18$) with a robustness effect that contributes a large upward drift.

To estimate the magnitude of this drift we assume, as in Broseta (1993a) and Crawford (1995), that $\bar{z}_{it} + \eta_{it}$ and z_{it} are approximately jointly normal for all i and t. Normality is a reasonable approximation because $\bar{z}_{it} + \eta_{it}$ and z_{it} are weighted sums of serially independent random variables. It makes the distributions of $(\bar{z}_{it} + \eta_{it})/\sigma_{pt}$ and z_{it}/σ_{zt} independent of t as well as i, so that $\mu_t \overset{t}{\equiv} \mu$ and $\nu_t \overset{t}{\equiv} \nu$. Given that $(\bar{z}_{it} + \eta_{it})/\sigma_{pt}$ and z_{it}/σ_{zt} are independent across i, μ and ν can be estimated [using (22) for ν] from D. Teichroew (1956 Table I) for any order statistic and $m \leq 20$. For VHBB's environment this yields $\mu = -0.069$ and $\nu = 0.665$, the respective means of the tenth and fifth largest [the $(n + 1)$st and the median of the n largest] of 18 i.i.d. standard normal variables.

Continuing to ignore the discreteness of effort, Proposition 6 can now be used to analyze the sources of the efficiency-enhancing effect of auctions in VHBB's experiment. Assume, as in Section III's econometric analysis, that $\varepsilon_t = \varepsilon$ for $t = 1, \dots$, and $\delta_t = \delta$ for $t = 0, \dots$, and set $\alpha_0 = \bar{\alpha}$ and $\alpha_t = 0$ for $t = 1, \dots$.[20] Proposition 6 then implies

$$(24)\quad Ex_{it} = \bar{\alpha} + \sum_{s=0}^{t} [\delta\beta_s + \gamma_s]$$

$$+ \delta\mu \sum_{s=0}^{t} \sigma_{ps} + \varepsilon\nu \sum_{s=0}^{t-1} \sigma_{zs}$$

and

$$(25)\quad Ey_t = \bar{\alpha} + \sum_{s=0}^{t} [\delta\beta_s + \gamma_s] + \delta\mu \sum_{s=0}^{t} \sigma_{ps}$$

$$+ \varepsilon\nu \sum_{s=0}^{t-1} \sigma_{zs} + \nu\sigma_{zt},$$

which express Ex_{it} and Ey_t as functions of μ, ν, and identified behavioral parameters. $\bar{\alpha}$ is the average initial level of players' beliefs; $\delta\beta_s$

[20] Setting $\alpha_t = 0$ involves no loss of generality because, as (26)–(27) make clear, of the four parameters in the constant terms of (18) and (19), $\alpha_t + \beta_t\delta_t + \gamma_t$, only δ_t and the combinations $\alpha_t + \beta_t$ and $(1 - \delta_t)\alpha_t + \gamma_t$ are identified. Setting $\alpha_t = 0$ replaces those combinations by β_t and γ_t, which now partly reflect the trends in beliefs in α_t.

TABLE 1—APPROXIMATIONS OF THE DYNAMICS FOR VHBB'S EXPERIMENT

t	$\bar{\alpha}$	$\delta\beta_t$	γ_t	$\delta\mu\sigma_{pt}$	$\varepsilon\nu\sigma_{zt-1}$	$\nu\sigma_{zt}$	Ex_{it}	Ey_t	$\overline{Ex}_{it}$
0	4.940	−0.050	0.297	−0.036	—	0.811	5.151	5.962	5.542
1	—	−0.060	0.358	−0.025	0.622	0.554	6.046	6.600	6.166
2	—	−0.068	0.400	−0.020	0.427	0.444	6.785	7.000	6.610
3	—	−0.073	0.432	−0.017	0.341	0.379	7.000	7.000	6.721
4	—	−0.077	0.458	−0.015	0.291	0.335	7.000	7.000	6.902

reflects the adjustment of players' bids for strategic uncertainty and any trends in beliefs; and γ_s combines the forward induction effect and trends in beliefs. $\delta\mu\sigma_{ps}$ is the effect of the dispersion of bids via q_t. $\varepsilon\nu\sigma_{zs}$ is the persistent component of the optimistic subjects/robustness effect, which incorporates the inertia of players' adjustments through ε. Finally, $\nu\sigma_{zt}$ in (25) is the direct effect of the dispersion of efforts on Ey_t. Part of this last effect is transient; after the period in which it first appears it is multiplied by ε and incorporated into the sum in the preceding term.

Table 1 presents approximations of Ex_{it} and Ey_t based on (24) and (25) and Section III's estimates of the behavioral parameters, reported in Table 2. They are computed under the parameter restrictions described above and simple intertemporal restrictions on β_t, γ_t, and the shock distributions imposed in estimation, described in Section III, subsection E. The second through sixth columns in Table 1 give the terms in (24) and (25) for periods 0–4, with the implied values of Ex_{it} and Ey_t and the average values of x_{it} and y_t in the eight trials of VHBB's experiment, $\overline{Ex}_{it}$ and $\overline{Ey}_t$, for comparison. Ex_{it} is computed by summing the second through fifth columns in row t and adding the result to Ex_{it-1}, truncating at 7.000 as necessary. (For Ex_{i0}, Ex_{it-1} is replaced by the estimated value of $\bar{\alpha}$, 4.940.) Ey_t is computed by adding the transient term $\nu\sigma_{zt}$ to Ex_{it} and truncating at 7.000.

Because some parameters are not precisely estimated, we view Table 1's results as illustrative. The approximations track the dynamics reasonably well, somewhat overestimating the rate at which Ex_{it} and $Ey_t \rightarrow 7$. They suggest that of the four components of Ex_{it} and Ey_t in our theoretical analysis, only two were important in VHBB's experiment: γ_t, forward induction, and $\varepsilon\nu\sigma_{zt-1}$, the optimistic subjects/robustness effect. Both favor efficiency, with roughly equal strength. In the three periods (0–2) it takes Ex_{it} and Ey_t to reach 6.785 and 7.000 from the initial level set by $\bar{\alpha} = 4.940$: γ_t contributes 1.049 to their increases, and $\varepsilon\nu\sigma_{zt-1}$ contributes 1.055. The transient term $\nu\sigma_{z2}$ contributes an additional 0.444 to the increase in Ey_t.[21]

Our methods can also be used to assess the likely effects of changes in the treatment variables, on the assumption that the behavioral parameters do not vary sharply with such changes. Consider replacing VHBB's median game with a nine-person minimum game, with all else unchanged. $\mu = -0.069$ as before because m and n are unchanged, but ν is now approximately the robustness effect in an 9-player game without auctions, in which payoffs and best responses are determined by the ninth largest (the minimum of the nine largest) of players' efforts. For this game Teichman's Table I yields $\nu = 0.069$, as the optimistic subjects effect is almost neutralized by the robustness effect of the minimum game. The tenfold fall in ν reduces $\varepsilon\nu\sigma_{zt-1}$ and $\nu\sigma_{zt}$ in Table 1 proportionately, yielding approxi-

[21] These contributions sum to more than 7 − 4.940 because the other two components contribute −0.253, and the approximated Ey_2 must be rounded down to 7.

TABLE 2—PARAMETER ESTIMATES

t	$\bar{\alpha}$	β_t	γ_t	λ_T	δ	ε	$\sigma^2_{\nu t}$	$\sigma^2_{\eta t}$	λ_V	ρ
0	4.940 (0.511)	−0.0949 (0.154)	0.2973 (0.221)	—	0.5287 (0.137)	—	1.4872 (0.383)	1.0000 —	—	0.5226 (0.217)
1	—	−0.1144 (0.177)	0.3583 (0.181)	−0.2689 (0.402)	0.5287 (0.137)	0.9348 (0.309)	0.6938 (0.132)	0.4665 (0.072)	1.0999 (0.223)	0.5226 (0.217)
2	—	−0.1276 (0.195)	0.3995 (0.156)	−0.2689 (0.402)	0.5287 (0.137)	0.9348 (0.309)	0.4442 (0.090)	0.2987 (0.073)	1.0999 (0.223)	0.5226 (0.217)
3	—	−0.1378 (0.211)	0.4317 (0.144)	−0.2689 (0.402)	0.5287 (0.137)	0.9348 (0.309)	0.3237 (0.075)	0.2177 (0.067)	1.0999 (0.223)	0.5226 (0.217)
4	—	−0.1464 (0.224)	0.4583 (0.144)	−0.2689 (0.402)	0.5287 (0.137)	0.9348 (0.309)	0.2533 (0.067)	0.1703 (0.061)	1.0999 (0.223)	0.5226 (0.217)
5	—	−0.1537 (0.236)	0.4814 (0.152)	−0.2689 (0.402)	0.5287 (0.137)	0.9348 (0.309)	0.2072 (0.061)	0.1393 (0.056)	1.0999 (0.223)	0.5226 (0.217)

Note: Asymptotic standard errors in parentheses; $\log L = -834.955$.

values for Ey_t of 5.235, 5.629, 6.031, 6.448, and 6.879 in periods 0–4.

These values are consistently higher than those observed in Weber's (1994) single trial in this environment, which yielded y_t of 3, 4, 1, 1, and 3 in periods 0–4. We suspect the discrepancy arises because the minimum is too far from the median to justify ignoring changes in the behavioral parameters, and because the approximations ignore the bounds on x_{it} and y_t. To predict outcomes accurately by our methods in this new environment would require empirical information about behavior in nearby environments.

F. *Analysis and Simulations with Discrete Efforts*

Most of our analysis extends to the model with discrete efforts $\hat{x}_{it}$ outlined in Section I, subsection D. The dynamics are still Markov, with state vector $\bar{\mathbf{x}}_t$ representing players' beliefs; the only difference is in how beliefs determine efforts and y_t. Adapting the arguments in Crawford (1995 Sec. 5) shows that Propositions 0 and 2–5 hold as stated, and Propositions 1 and 6 hold approximately when the grid of feasible efforts is fine enough.

The prior probability distributions of the dynamics and limiting outcome can be estimated precisely for the model with discrete efforts only by repeated simulation. As a check on the approximations we computed frequency distributions for 5,000 simulation runs with the estimated parameters reported in Table 2, under normality assumptions described in Section III, subsection A. The simulation results are reported in Tables 3 and 4 and Figure 1. Comparing the simulations with the approximations in Section II, subsection E, suggests that the approximations overstate the rate of increase of $\hat{x}_{it}$ and y_t because they ignore their upper bounds, which cut off the upper tails of the normal distributions on which our estimates of μ_t and ν_t are based. Although this problem will probably arise whenever either an upper or a lower bound comes disproportionately into play, the analysis of Crawford (1995) suggests that our approximation technique will otherwise yield reasonably accurate results.[22]

[22] Crawford and Broseta (1995 Figures II–IV) report simulation estimates of Eq_t and Ey_t for alternative environments in which $f(\cdot)$ ("j" in Section II, subsection D) varies from the fifth to the third or seventh order statistic, n varies from 9 to 7 or 11, and m varies from 18 to 16 or 20, holding other treatment variables constant at VHBB's values. These results suggest that, at least for moderate

TABLE 3—MEAN VALUES OF q_t AND y_t ($t = 0, \ldots, 9$)

t	Actual mean q_t, y_t	Predicted mean q_t, y_t
0	4.775, 5.750	4.782, 5.648
1	5.650, 6.375	5.454, 6.207
2	6.213, 6.625	6.465, 6.864
3	6.450, 6.875	6.768, 6.963
4	6.700, 6.875	6.921, 6.993
5	6.775, 7.000	6.979, 6.998
6	6.850, 7.000	6.996, 6.999
7	6.838, 7.000	6.999, 7.000
8	6.913, 7.000	7.000, 7.000
9	6.988, 7.000	7.000, 7.000

III. Econometric Specification and Estimation

This section provides a general econometric framework for estimating the behavioral parameters and reports estimates and tests for VHBB's (1993) experimental data.

A. *The Model*

We continue to assume that players' bids, p_{it}, the market-clearing price, q_t, and their beliefs, x_{it}, are continuously variable, but we now assume that players' efforts are determined by an ordered probit model of discrete choice, described in Section III, subsection B.

Substituting $\bar{x}_{it}$ from (9) into (7)–(8) yields

$$p_{it} = \beta_t + \alpha_t + \varepsilon_t y_{t-1} + (1 - \varepsilon_t)x_{it-1} + \zeta_{it} + \eta_{it} \tag{26}$$

and

$$x_{it} = \gamma_t + (1 - \delta_t)\alpha_t + \delta_t q_t + (1 - \delta_t)\varepsilon_t y_{t-1} + (1 - \delta_t)(1 - \varepsilon_t)x_{it-1} + (1 - \delta_t)\zeta_{it} + \theta_{it} \tag{27}$$

for $t = 1, \ldots,$ and $i = 1, \ldots, m$, with initial conditions given by setting presample values to 0. For simplicity we write $\tau_{it} \equiv \zeta_{it} + \eta_{it}$ and $\nu_{it} \equiv (1 - \delta_t)\zeta_{it} + \theta_{it}$. We maintain the following distributional assumptions. Conditional on $\mathbf{I}_t$, the history of p_{is} and x_{is} through time t (up to and including y_t), the vector of innovations (τ_{it}, ν_{it}) is jointly i.i.d. across i, serially uncorrelated, and has a bivariate normal distribution with means 0, variances $\sigma^2_{\tau t} \equiv \sigma^2_{\zeta t} + \sigma^2_{\eta t}$ and $\sigma^2_{\nu t} \equiv (1 - \delta_t)^2\sigma^2_{\zeta t} + \sigma^2_{\theta t}$, and covariance $(1 - \delta_t)\sigma^2_{\zeta t}$.

The model defined by (26)–(27) determines all m players' current bids and beliefs, as functions of the observed values of y_{t-1} and q_t, and players' past beliefs x_{it-1} (but not their past efforts). For any given t, player i plays the coordination game if and only if (barring ties, which have zero probability) $p_{it} \geq q_t$. This yields two regimes for x_{it}: one in which $p_{it} < q_t$ and x_{it} is completely censored; and one in which $p_{it} \geq q_t$ and x_{it} is not directly observed but acts as a latent variable determining player i's observed effort. Thus, sample separation is determined endogenously by whether or not $p_{it} < q_t$.[23]

The resulting model raises three econometric issues: discrete choice, endogenous sample selection, and latent explanatory variables. We discuss these in Section III, subsections B–D, constructing the likelihood function as we go. We then discuss specification and estimation issues and report estimates in Section III, subsection E.

changes, all three of these treatment variables have significant effects, close to those suggested by our approximations.

[23] We assume that the p_{it} are unobserved, and to make the computations tractable we treat q_t as exogenous. These are reasonable approximations in VHBB's treatment

TABLE 4—DYNAMICS OF y_t ($t = 0, \ldots, 6$)

	y_0	y_1	y_2	y_3	y_4	y_5	y_6
7	0.125	0.625	0.625	0.875	0.875	1.000	1.000
	0.055	0.329	0.654	0.871	0.964	0.994	0.999
6	0.500	0.125	0.375	0.125	0.125	0.000	0.000
	0.550	0.551	0.315	0.122	0.036	0.006	0.001
5	0.375	0.250	0.000	0.000	0.000	0.000	0.000
	0.382	0.117	0.031	0.007	0.000	0.000	0.000
4	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.013	0.002	0.000	0.000	0.000	0.000	0.000
≤3	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Note: Actual frequency distributions in first lines, predicted distributions in second lines.

B. *Discrete Choice*

Because players' efforts are naturally ordered by their payoff implications, we use an ordered probit discrete-choice model, as in Daniel McFadden (1984), in which the x_{it} determine players' bids and efforts. We partition the state space of x_{it}—the extended real line—into response regions C_k ($k = 1, \ldots, K$), where C_k is the interval $(a_{k-1}, a_k]$, and $a_0 = -\infty < a_1 < \cdots < a_K = +\infty$. We then set $\hat{x}_{it} = k$ if and only if $x_{it} \in C_k$. Under our normality assumption, player i's choice probabilities at time t, conditional on $\mathbf{I}_{t-1}$, can be written

$$\Pr[x_{it} \in C_k \mid \mathbf{I}_{t-1}] = \int_{C_k} f_x(x_{it})\, dx_{it} \tag{28}$$

$$= \Phi[(a_k - E_{t-1}x_{it})/\sigma_{\nu_t}]$$

$$- \Phi[(a_{k-1} - E_{t-1}x_{it})/\sigma_{\nu_t}],$$

$$k = 1, \ldots, K,$$

where $f_x(\cdot)$ and $\Phi(\cdot)$ denote the marginal density of x_{it} and the standard normal cumulative distribution function, respectively.

It is commonly assumed that the thresholds a_k are independent of i and t. We make the stronger assumption that $a_k = k + {}^1/_2$ for $k = 1, \ldots, 6$, so that $\hat{x}_{it}$ is determined simply by rounding x_{it} to the nearest integer in the set $\{1, \ldots, 7\}$.[24]

C. *Endogenous Sampling*

Because the subsample of n players whose efforts are observed each period is endogenously determined, and τ_{it} and ν_{it} are generally correlated for any given player, standard maximum-likelihood estimates of the parameters in (26)–(27) will suffer from sample selection bias unless corrective action is taken.[25]

Let Γ_t denote the set $\{i \mid p_{it} \geq q_t\}$ of players who (barring ties) play the coordination game at time t, which VHBB recorded along with the other data from their experiment. We

[24] There is no loss of generality in identifying the scales on which $\hat{x}_{it}$ and x_{it} are measured. The equal spacing of thresholds is restrictive, but it is optimal in VHBB's median game when players are risk neutral and x_{it} is player i's estimate of $E[y_t \mid \mathbf{I}_{t-1}]$ (Crawford 1995 footnote 27 p. 127), and fixing the a_k greatly reduces the number of parameters to be estimated and identifies the conditional variance of ν_{it}.

[25] Various corrective methods have been proposed in the literature. See for example G. S. Maddala (1983 Ch. 9) for limited dependent variable models and Charles Manski and McFadden (1981) or Takeshi Amemiya (1985 Ch. 9) for qualitative response models.

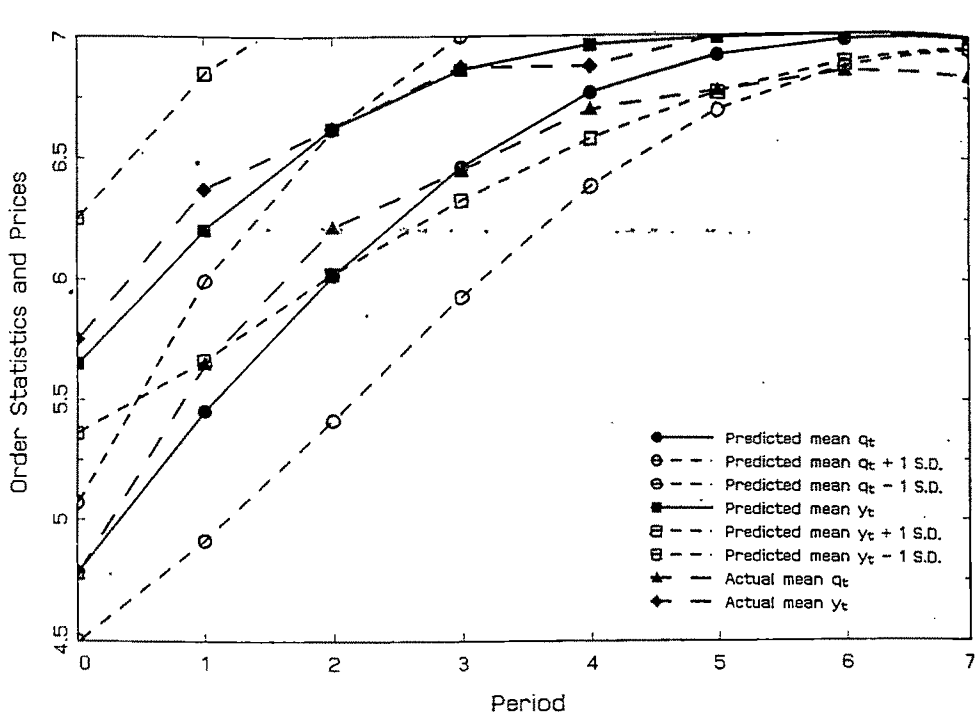


FIGURE 1. SIMULATED DYNAMICS

propose a full-sample estimation procedure, in which (26)–(27) are estimated simultaneously using all available information on subjects' decisions. For each $k = 1, \ldots, 7$, define the dummy variable $D^k_{it} \equiv 1$ when $\hat{x}_{it} = k$ and 0 otherwise. Let $\boldsymbol{\omega}_t$ denote the vector of parameters of interest at time t. Conditional on $\mathbf{I}_{t-1}$, the contribution of the (it)th observation to the likelihood function can then be written

$$(29)\quad L_{it}(\boldsymbol{\omega}_t) = 1_{\{i \notin \Gamma_t\}} \int_{-\infty}^{q_t} f_p(p_{it})\,dp_{it} + 1_{\{i \in \Gamma_t\}} \sum_{k=1}^{7} \left[\int_{C_k} \int_{q_t}^{\infty} f(x_{it}, p_{it})\,dx_{it}\,dp_{it} \right]^{D^k_{it}},$$

where $1_{\mathcal{A}}$ is the indicator function of the set $\mathcal{A}$, and $f(\cdot)$ and $f_p(\cdot)$ denote the joint conditional density of (x_{it}, p_{it}) and the marginal density of p_{it}, respectively. The two regimes in (29) reflect observations in which subject i does not play and we learn only that $p_{it} < q_t$, and those in which subject i plays and we learn both that $p_{it} \geq q_t$ and that $x_{it} \in C_k$ when $\hat{x}_{it} = k$. In this respect our model resembles differential response models for analyzing cross-section data from surveys with voluntary participation (Stephen Pudney, 1989 Sec. 2...) or James Heckman's (1974) model of labor supply, with the added complications of the dynamic structure and the double integral in (29).

This estimation technique makes great demands on the sample. Both the sample selection equation, (26), and the equation describing the adjustment of beliefs, (27), must be estimated from observations of $\hat{x}_{it}$, y_t, and q_t; and q_t, for which we have only eight observations per period, plays a crucial role in identifying the parameters. Muthén and Joreskog's results, reported in Maddala (198..., p. 267), suggest that we may not obtain very precise estimates of the parameters $\beta_t + \alpha_t$ and $\sigma^2_{\tau_t}$ in (26).

D. Latent Explanatory Variables

Standard estimation techniques, such as two-stage probit and Tobit models (D... Grether and Maddala [1982] and Maddala [1983 Ch. 9]), deal with the latent vari...

x_{it-1} on the right-hand sides of (26) and (27) by estimating the reduced form of the model. Here, however, the error terms of the reduced form are serially correlated, so that full maximum-likelihood estimation would involve calculating the T-fold integral of the joint distribution of x_{it} and p_{it} ($i = 1, \ldots, m$; $t = 0, \ldots, T$). This is computationally intractable even for simple types of serial correlation, and our model already involves additional difficulties due to endogenous sampling and a nonstationary error structure.

We address these problems by using the reduced forms of (26)–(27) to maximize the likelihood function $L(\omega) \equiv \Pi_{i,t} L_{it}(\omega_t)$, with $\omega \equiv (\omega_0, \ldots, \omega_T)$ and $L_{it}(\cdot)$ as in (29), with respect to the parameters of interest. Thus we ignore the serial correlation, which will result in inefficient parameter estimates, while taking the heteroskedasticity of the error terms fully into account as described below. We are not aware of any consistency or asymptotic normality results for maximum-likelihood estimators in models like ours. Thus our estimates of the behavioral parameters and standard errors must be interpreted with caution.

E. *Empirical Specification and Estimation*

As explained in Section I, subsection C, there is a danger of overfitting unless we rule out ad hoc time variation in the behavioral parameters. We therefore impose simplifying intertemporal restrictions on the parameters, which allow an adequate econometric description of behavior. As in Section II, subsection E, we set $\varepsilon_t = \varepsilon$ for $t = 1, \ldots$ and $\delta_t = \delta$ for $t = 0, \ldots$. We also set $\alpha_0 = \bar{\alpha}$ and $\alpha_t = 0$ for $t = 1, \ldots$, so that $\alpha_t + \beta_t = \beta_t$ and $(1 - \delta)\alpha_t + \gamma_t = \gamma_t$ for $t = 1, \ldots$; this involves no loss of generality because only $\alpha_t + \beta_t$ and $(1 - \delta)\alpha_t + \gamma_t$ are identified. We further assume that $\beta_t = \beta_0(1 + t)^{-\lambda_T}$ and $\gamma_t = \gamma_0(1 + t)^{-\lambda_T}$ for $t = 1, \ldots$. Finally, following Broseta (1993b) and Crawford (1995), we impose the following intertemporal constraints on the variance-covariance matrix of the innovations τ_{it} and ν_{it} in (26)–(27): $\sigma^2_{\tau_t} = \sigma^2_{\tau_0}(1 + t)^{-\lambda_\nu}$, $\sigma^2_{\nu_t} = \sigma^2_{\nu_0}$ [illegible], and $\text{Cov}(\tau_{it}, \nu_{it}) = \text{Cov}(\tau_{i0}, \nu_{i0})$ [illegible], so that the correlation coefficient [illegible] $= \rho$ for $t = 0, \ldots$.

Under these assumptions, backward substitution for x_{it-1} in (26)–(27) yields

$$(30)\quad p_{it} = \beta_0(1 + t)^{-\lambda_T} + \gamma_0(1 - \varepsilon) \times \sum_{k=1}^{t-1} [(1 - \delta)(1 - \varepsilon)]^{k-1}(t - k + 1)^{-\lambda_T} + (1 - \varepsilon)^t(1 - \delta)^{t-1}[\gamma_0 + (1 - \delta)\bar{\alpha}] + \varepsilon \sum_{k=1}^{t} [(1 - \delta)(1 - \varepsilon)]^{k-1} y_{t-k} + \delta(1 - \varepsilon) \times \sum_{k=1}^{t} [(1 - \delta)(1 - \varepsilon)]^{k-1} q_{t-k} + \tau_{it} + (1 - \varepsilon) \sum_{k=1}^{t} [(1 - \delta)(1 - \varepsilon)]^{k-1} \nu_{it-k}$$

and

$$(31)\quad x_{it} = \gamma_0 \sum_{k=0}^{t-1} [(1 - \delta)(1 - \varepsilon)]^k \times (t - k + 1)^{-\lambda_T} + [(1 - \varepsilon)(1 - \delta)]^t[\gamma_0 + (1 - \delta)\bar{\alpha}] + \varepsilon(1 - \delta) \times \sum_{k=1}^{t} [(1 - \delta)(1 - \varepsilon)]^{k-1} y_{t-k} + \delta \sum_{k=0}^{t} [(1 - \delta)(1 - \varepsilon)]^k q_{t-k} + \sum_{k=0}^{t} [(1 - \delta)(1 - \varepsilon)]^k \nu_{it-k}.$$

Letting $e_{it} \equiv \tau_{it} + (1 - \varepsilon) \Sigma_{k=1}^{t} [(1 - \delta)(1 - \varepsilon)]^{k-1}\nu_{it-k}$ and $u_{it} \equiv \Sigma_{k=0}^{t} [(1 - \delta)(1 - \varepsilon)]^k \nu_{it-k}$ represent the error terms in (30) and (31), it is clear that the e_{it} and u_{it} are serially correlated for any given i, but

uncorrelated across i. The relevant components of the variance-covariance matrix are now given, for $t = 0, \ldots,$ by

$$(32)\quad \mathrm{Var}(u_{it}) = \sigma^2_{\nu_0} \sum_{k=0}^{t} [(1-\delta)(1-\varepsilon)]^{2k} \times (t-k+1)^{-\lambda_V},$$

$$(33)\quad \mathrm{Var}(e_{it}) = \sigma^2_{\tau_0}(1+t)^{-\lambda_V} + (1-\varepsilon)^2\sigma^2_{\nu_0} \times \sum_{k=1}^{t} [(1-\delta)(1-\varepsilon)]^{2(k-1)} \times (t-k+1)^{-\lambda_V},$$

and

$$(34)\quad \mathrm{Cov}(e_{it}, u_{it}) = \mathrm{Cov}(\tau_{i0}, \nu_{i0})(1+t)^{-\lambda_V} + (1-\varepsilon)\sigma^2_{\nu_0} \times \sum_{k=1}^{t} [(1-\delta)(1-\varepsilon)]^{2k-1} \times (t-k+1)^{-\lambda_V}.$$

For each period there are eight observations of q_t and y_t and, counting both regimes in (29), 144 of $\hat{x}_{it}$. The model defined by (28)–(34) was estimated by maximum likelihood, using the data from the first six periods of VHBB's eight trials.[26] There are ten parameters of interest: $\bar{\alpha}$, β_0, γ_0, λ_T, δ, ε, $\sigma^2_{\tau_0}$, $\sigma^2_{\nu_0}$, λ_V, and ρ.[27] All ten are theoretically identified. However, $\sigma^2_{\tau_0}$ is only weakly identified, because [by (... when the p_{it} are unobserved the only observable consequence of an increase in their variance is an increase in the variance of q_t. In practice the estimated $\sigma^2_{\tau_0}$ "blows up" because when $n/m = 0.5$, as in VHBB's design, it makes the fit in the sample selection equation (26) approximately independent of the estimated β_0 and ε, which can then be chosen to maximize the rest of the likelihood function. We deal with this problem by estimating the model under the normalization $\sigma^2_{\tau_0} \equiv 1$, which is arbitrary but unavoidably so when the p_{it} are unobserved. Fortunately, as just explained, the level of the $\sigma^2_{\tau_t}$ (as opposed to their time trend λ_V, which is precisely estimated) has little effect on the model's implications for observable variables.

Table 2 reports the estimated parameters with asymptotic standard errors in parentheses. The estimates confirm that the model gives an adequate econometric description of VHBB's subjects' behavior. With the exception of λ_T, the estimate of which is slightly and insignificantly less than zero, they satisfy the restrictions suggested by the theory.[28] The estimated adjustment coefficients δ and ε are plausible, though the latter, at 0.93, is much higher than the 0.58 estimate for the analogous parameter in the analogous treatment with auctions (Crawford, 1995 Table I) and somewhat higher than our prior. The estimated [illegible] are positive and quite large, also somewhat higher than our prior; together with our estimate of δ they imply a large forward induction effect, raising efforts nearly half a unit above the value suggested by q_t and prior beliefs, on average. The estimated β_t are negative as expected, small (recalling that bids are scaled like efforts) and insignificantly different from 0. The estimated $\sigma^2_{\nu_t}$ are significantly positive, so that constraints ruling out strategic uncertainty are strongly rejected. The estimated

[26] Although the trials lasted 10–15 periods, convergence was so rapid that there was little sample variation after period 6 and none after period 7. We excluded the data in every period for subject 1 in trial 10, who paid a high price to play the coordination game in period 3 and then played the lowest effort. Preliminary estimates revealed that this subject's behavior had a large impact on the estimated $\sigma^2_{\nu_0}$. The program code and the results of intermediate computations are available on request.

[27] Ten may seem like a large number, but Section II's analysis shows that to explain VHBB's result the model must describe the dynamics of a heterogeneous population of learning rules. Our estimates strongly suggest that no simpler model can do this, and if tighter estimates are desired further experimental data can be produced at will.

[28] λ_T is naturally positive because it represents the rate of *decline* of β_t and γ_t, which should approach zero over time as strategic uncertainty is eliminated by experience. Recall that α_t has been normalized to zero, so that β_t and γ_t include any exogenous trends in beliefs that show up in players' bids (β_t) and/or efforts (γ_t).

relation coefficient ρ is significantly positive, as expected because ζ_{it} enters both τ_{it} and ν_{it}, with positive weights. Finally, the estimated λ_V is significantly positive, confirming that there were substantial learning effects, and greater than one (though insignificantly so), so that Proposition 2's variance conditions for convergence are satisfied.[29]

We estimated the model's implications for the prior probability distributions of the dynamics and limiting outcomes by repeated simulation, as described in Section II, subsection F, taking the discreteness and boundedness of efforts fully into account. Table 3 and Figure 1 report the actual and predicted mean values of q_t and y_t, and Table 4 reports the predicted distributions of y_t in the second lines of each cell, with the actual frequency distributions in the first lines for comparison. Figure 1 illustrates the dynamic interplay between q_t and y_t that underlies the efficiency-enhancing effect, in which q_t exerts upward pressure on the distributions of the x_{it} and y_t via forward induction, y_t then exerts upward pressure on the distributions of the $\bar{x}_{it+1}$, the p_{it+1}, and q_{t+1}, and so on.[30]

The actual and predicted means and distributions seem very close, with the single exception of the distribution of y_1. Their closeness is difficult to judge by eye, because the frequency distributions for eight trials must be coarser than the predicted distributions. However, χ^2 tests of goodness of fit, conducted for the distribution of y_t each period, never come close to rejecting the hypothesis that the actual frequencies were drawn from the predicted distributions.[31] The relevant χ^2 statistics, each with six degrees of freedom, are 0.854 with p-value 0.991 in period 0; 5.999 with p-value 0.423 in period 1; 1.826 with p-value 0.935 in period 4; and 0.350, 0.057, 0.048, and 0.008 in periods 2, 3, 5, and 6. These results suggest that the model closely reproduces the dynamics of subjects' interactions.

IV. Conclusion

In this paper we propose, estimate, and analyze a model of the learning dynamics in a recent experiment by Van Huyck et al. (1993), whose subjects repeatedly played a coordination game with seven Pareto-ranked equilibria, with the right to play auctioned each period in a larger group. The auctions had a striking efficiency-enhancing effect, in that subjects invariably bid the market-clearing price to a level recoverable only in the efficient equilibrium and then converged to that equilibrium, although subjects who played the same coordination game without auctions always converged to inefficient equilibria. The efficiency-enhancing effect of auctioning the right to play suggests a new and potentially important way in which competition may promote efficiency. And its effectiveness in focusing subjects' expectations on desirable equilibria may eventually help to unify our understanding of preplay communication, via "cheap talk" as well as costly signaling.

VHBB's contribution is a good example of how carefully designed and conducted experiments can change the way we think about important economic problems. Although it conveys a powerful impression by itself, its power to inform analysis can only be fully realized by a theory that identifies the mechanism behind the efficiency-enhancing effect and provides a firm basis for generalization to other laboratory or field environments. Working with a model that is flexible enough to nest alternative explanations, as in our earlier analyses of equilibrium selection in VHBB's (1990, 1991) experiments in Broseta (1993a, 1993b) and Crawford (1995), we find that explaining VHBB's (1993) result requires a model of stochastic, history-dependent learning dynamics, in which interactions between

[29] Because $\lambda_T < 0$, Proposition 2's conditions on β_s and γ_s are violated. As explained in Section II, subsection C, convergence occurs despite these violations because the are bounded.

[30] The upward pressure on the x_{it} and y_t from q_t comes mostly from an estimated $\gamma_t > 0$. q_{t+1} is not generally higher than y_t because the estimated $\beta_t < 0$ and $\varepsilon < 1$.

[31] These tests are not independent across periods because of the history dependence; we offer them only as a way to gauge how closely the model reproduces the data.

nonnegligible amounts of strategic uncertainty and the learning dynamics have a persistent effect on the efficiency of the limiting outcome. Our analysis quantifies this effect, relating it to the treatment variables (the numbers of players in the coordination game and the auction, and the order statistic that determines the robustness of desirable equilibria) in a way that makes it possible to assess its likely magnitude in other environments. We find that auctions can be expected to enhance efficiency to some extent in a wide range of environments that share the basic ingredients of the mechanism we identify—our optimistic subjects, forward induction, and robustness effects. However, in other environments this effect may be too small to assure fully efficient coordination.

Our methodology offers some more general insights about the analysis of learning and equilibrium selection. Because the extent of strategic uncertainty cannot usefully be explained by theory alone, its persistent effect on the limiting outcome gives the analysis an empirical component, which we describe via exogenous behavioral parameters in our specification of players' learning rules. Much of our analysis is independent of the precise values of these parameters, but they do affect the quantitative results. We close the model, when necessary, by estimating them using the experimental data. This dependence on empirical parameters and the need to analyze stochastic, history-dependent learning dynamics are likely to be encountered in realistic models of equilibrium selection in other environments. Although our analysis does not immediately generalize beyond VHBB's environments, we hope it shows that such analyses need not be intractable, and that the theoretical and empirical understanding needed to address economic questions involving equilibrium selection and coordination may be achievable in other settings.

Appendix

PROOF OF PROPOSITION 0:

Under the stated conditions $\eta_{it} \equiv \theta_{it} \equiv \zeta_{it} \equiv 0$ for $t = T, \dots$. (11) reduces to

$$\text{(A1)} \quad \bar{x}_{it+1} = \delta_t g(\bar{x}_{1t}, \dots, \bar{x}_{mt}) + \varepsilon_{t+1}(1-\delta_t) \times h(\bar{x}_{1t}, \dots, \bar{x}_{mt}; \bar{x}_{1t}, \dots, \bar{x}_{mt}) + (1-\varepsilon_{t+1})(1-\delta_t)\bar{x}_{it}.$$

Because δ_t, ε_{t+1}, $g(\bar{x}_{1t}, \dots, \bar{x}_{mt})$, and $h(\bar{x}_{1t}, \dots, \bar{x}_{mt}; \bar{x}_{1t}, \dots, \bar{x}_{mt})$ are the same for all players, and $0 \leq (1-\varepsilon_{t+1})(1-\delta_t) < 1$, (A1) leaves the order of the $\bar{x}_{it}$ (weakly) unchanged over time. It follows that there exist players j and k, such that $g(\bar{x}_{1t}, \dots, \bar{x}_{mt}) = \bar{x}_{jt}$ and $h(\bar{x}_{1t}, \dots, \bar{x}_{mt}; \bar{x}_{1t}, \dots, \bar{x}_{mt}) = \bar{x}_{kt}$ for all $t = T, \dots$. (A1) then implies

$$\text{(A2)} \quad \bar{x}_{jt+1} = \delta_t \bar{x}_{jt} + \varepsilon_{t+1}(1-\delta_t)\bar{x}_{kt} + (1-\varepsilon_{t+1})(1-\delta_t)\bar{x}_{jt} = [1-\varepsilon_{t+1}(1-\delta_t)]\bar{x}_{jt} + \varepsilon_{t+1}(1-\delta_t)\bar{x}_{kt}$$

and

$$\text{(A3)} \quad \bar{x}_{kt+1} = \delta_t \bar{x}_{jt} + \varepsilon_{t+1}(1-\delta_t)\bar{x}_{kt} + (1-\varepsilon_{t+1})(1-\delta_t)\bar{x}_{kt} = \delta_t \bar{x}_{jt} + (1-\delta_t)\bar{x}_{kt}.$$

When $0 < \delta_t, \varepsilon_t \leq 1$, (A2) and (A3) imply that $\bar{x}_{jt}$ and $\bar{x}_{kt}$ converge to a common value weakly between $\bar{x}_{jT}$ and $\bar{x}_{kT}$, which is completely determined by $\bar{x}_{jT}$, $\bar{x}_{kT}$, the δ_t, and the ε_t. It is clear from (A1) and (8) that the other $\bar{x}_{it}$ and x_{it} converge to the same value, and from (6)–(8), (11), (A2), and (A3) that if $T = 0$, $\bar{x}_{it} \equiv x_{it} \equiv \alpha_0$ for all i and $t = 0, \dots$.

PROOF OF PROPOSITION 1:

As in Crawford (1995 Proposition 1), the proof is immediate by induction once the solution has been found. Here we give a method of constructing the solution, which is informative. Substituting from (1), (3), (6)–(9) yields

$$\text{(A4)}\quad y_t - y_{t-1} \equiv h(x_{1t}, \dots, x_{mt}; p_{1t}, \dots, p_{mt}) - h(x_{1t-1}, \dots, x_{mt-1}; p_{1t-1}, \dots, p_{mt-1})$$
$$= \alpha_t + \beta_t\delta_t + \gamma_t + \delta_t g[(1-\varepsilon_t)x_{1t-1} + \zeta_{1t} + \eta_{1t}, \dots, (1-\varepsilon_t)x_{mt-1} + \zeta_{mt} + \eta_{mt}]$$
$$+ h[(1-\delta_t)(1-\varepsilon_t)x_{1t-1} + (1-\delta_t)\zeta_{1t} + \theta_{1t}, \dots, (1-\delta_t)(1-\varepsilon_t)x_{mt-1} + (1-\delta_t)\zeta_{mt} + \theta_{mt}; p_{1t}, \dots, p_{mt}]$$
$$- (1-\varepsilon_t)h[x_{1t-1}, \dots, x_{mt-1}; p_{1t-1}, \dots, p_{mt-1}].$$

It is clear from (12) that z_{it} is the idiosyncratic component of x_{it}, from (9) and (15) that $\bar{z}_{it}$ is the idiosyncratic component of $\bar{x}_{it}$, and from (7) that $\bar{z}_{it} + \eta_{it}$ is the idiosyncratic component of p_{it}. The second part of (15) is immediate from (9), and (8)–(9) imply that

$$\text{(A5)}\quad z_{it} = (1-\delta_t)(1-\varepsilon_t)z_{it-1} + (1-\delta_t)\zeta_{it} + \theta_{it},$$

which yields the first part of (15) by successive substitution. Using (2), (4)–(5), (18), and (A5), (A4) can be rewritten

$$\text{(A6)}\quad y_t - y_{t-1} = \alpha_t + \beta_t\delta_t + \gamma_t + \delta_t g(\bar{z}_{1t} + \eta_{1t}, \dots, \bar{z}_{mt} + \eta_{mt})$$
$$+ h(z_{1t}, \dots, z_{mt}; \bar{z}_{1t} + \eta_{1t}, \dots, \bar{z}_{mt} + \eta_{mt})$$
$$- (1-\varepsilon_t)h(z_{1t-1}, \dots, z_{mt-1}; \bar{z}_{1t-1} + \eta_{1t-1}, \dots, \bar{z}_{mt-1} + \eta_{mt-1}).$$

Summing (A6) and substituting from (14) yields (13). To derive (12) from (13), note that

$$\text{(A7)}\quad x_{it} - y_t \equiv x_{it} - h(x_{1t}, \dots, x_{mt}; p_{1t}, \dots, p_{mt})$$
$$\equiv z_{it} - h(z_{1t}, \dots, z_{mt}; \bar{z}_{1t} + \eta_{1t}, \dots, \bar{z}_{mt} + \eta_{mt})$$
$$\equiv z_{it} - h_t.$$

This completes the proof.

PROOF OF PROPOSITION 2:

The proof follows the martingale convergence arguments of Nevel'son and Has'minskii (1973 Theorem 2.7.3). Substituting (1), (3), and (7)–(9) into (26)–(27) shows that $\mathbf{x}_t$ can be taken as the state vector instead of $\bar{\mathbf{x}}_t$, which is convenient because it is the x_{it}, not the $\bar{x}_{it}$, that are directly affected by the bounds and by the discreteness of effort. Define the Lyapunov function $V_t \equiv \Sigma_{i,j}(x_{it} - x_{jt})^2$, where the summation is taken over all $i, j = 1, \dots, m$.[32] Clearly, $V_t \geq 0$ for all $\mathbf{x}_t$, with $V_t = 0$ if and only if $x_{it} = x_{jt}$ for all i and j. Substituting from (27) and simplifying yields

$$\text{(A8)}\quad V_t = \sum_{i,j}[(1-\varepsilon_t)(1-\delta_t)(x_{it-1} - x_{jt-1}) + (1-\delta_t)(\zeta_{it} - \zeta_{jt}) + \theta_{it} - \theta_{jt}]^2$$
$$= \sum_{i,j}[(1-\varepsilon_t)^2(1-\delta_t)^2(x_{it-1} - x_{jt-1})^2 + 2(1-\varepsilon_t)(1-\delta_t)(x_{it-1} - x_{jt-1}) \times \{(1-\delta_t)(\zeta_{it} - \zeta_{jt}) + \theta_{it} - \theta_{jt}\} + \{(1-\delta_t)(\zeta_{it} - \zeta_{jt}) + \theta_{it} - \theta_{jt}\}^2].$$

[32] Nevel'son and Has'minskii assume that $V_t \to \infty$ as $\|\mathbf{x}_t\| \to \infty$, but they use this condition only to ensure that solution paths are bounded; we assume boundedness directly.

Taking expectations in (A8) conditional on $\mathbf{x}_{t-1}$ then yields

$$\text{(A9)}\quad E_{t-1}(V_t|\mathbf{x}_{t-1})$$

$$= (1-\varepsilon_t)^2(1-\delta_t)^2 \sum_{i,j} (x_{it-1} - x_{jt-1})^2$$

$$+ 2(1-\varepsilon_t)(1-\delta_t)$$

$$\times \sum_{i,j} E\{(1-\delta_t)(\zeta_{it}-\zeta_{jt}) + \theta_{it} - \theta_{jt}\}$$

$$\times (x_{it-1} - x_{jt-1})$$

$$+ \sum_{i,j} E\{(1-\delta_t)(\zeta_{it}-\zeta_{jt}) + \theta_{it} - \theta_{jt}\}^2.$$

The first term on the right-hand side of (A9) is plainly bounded below V_{t-1} for all $\mathbf{x}_{t-1}$ outside any given neighborhood of the set for which $V_{t-1} = 0$. Without the bounds the second term equals 0, and the third term converges to 0 with probability 1 because the finiteness of $\Sigma_{s=0}^{\infty}\sigma_{\zeta s}^2$ and $\Sigma_{s=0}^{\infty}\sigma_{\theta s}^2$ implies that $(1-\delta_t)^2\sigma_{\zeta t}^2 + \sigma_{\theta t}^2 \to 0$, which implies that $E[(1-\delta_t)\zeta_{it} + \theta_{it}]^2 \to 0$ with probability 1. Thus, without the bounds $\{V_t\}$ eventually becomes a nonnegative supermartingale, so that $V_t \to 0$ with probability 1. Because the bounds can never increase $E[(1-\delta_t)\zeta_{it} + \theta_{it}]^2$ or [by (A8)] $E_{t-1}(V_t|\mathbf{x}_{t-1})$, this is also true for the bounded version of the $\{V_t\}$ process. In either case, it follows that for all i and j, $(x_{it} - x_{jt}) \to 0$ with probability 1. Using (9), the strong law of large numbers, and the continuity of $h(\cdot)$ then shows that $(y_t - x_{it})$ and $(\bar{x}_{it} - \bar{x}_{jt}) \to 0$ with probability 1. y_t, $\bar{x}_{it}$, and x_{it} must then converge to a common limit because $\bar{\mathbf{x}}_t$ and $\mathbf{x}_t$ cannot (with positive probability) return infinitely often to a given point at which $\bar{x}_{it} \neq \bar{x}_{jt}$ or $x_{it} \neq x_{jt}$ for some i and j; and $\bar{\mathbf{x}}_t$ or $\mathbf{x}_t$ cannot (with positive probability) oscillate infinitely often between distinct points at which $\bar{x}_{it} = \bar{x}_{jt}$ or $x_{it} = x_{jt}$ for all i and j, because $\Sigma_{s=0}^{t}\alpha_s$, $\Sigma_{s=0}^{t}\beta_s$, $\Sigma_{s=0}^{t}\gamma_s$, $\Sigma_{s=0}^{t}\theta_{is}$, and $\Sigma_{s=0}^{t}\zeta_{is}$ also converge to finite limits (the last two with probability 1). The convergence of p_{it} and q_t to the same limit follows from the continuity of $g(\cdot)$ and the facts that $\beta \to 0$ as $t \to \infty$ and $\Sigma_{s=0}^{t}\eta_{is}$ converges with probability one.

REFERENCES

Amemiya, Takeshi. *Advanced econometrics*. Cambridge, MA: Harvard University Press, 1985.

Ben-Porath, Elchanen and Dekel, Eddie. "Signaling Future Actions and the Potential for Sacrifice." *Journal of Economic Theory*, June 1992, *57*(1), pp. 36–51.

Broseta, Bruno. "Strategic Uncertainty and Learning in Coordination Games." Department of Economics Discussion Paper No. 93-34, University of California, San Diego, August 1993a.

———. "Estimation of a Game-Theoretic Model of Learning: An Autoregressive Conditional Heteroskedasticity Approach." Department of Economics Discussion Paper No. 93-35, University of California, San Diego, August 1993b.

Cachon, Gerard and Camerer, Colin. "Loss Avoidance and Forward Induction in Experimental Coordination Games." *Quarterly Journal of Economics*, February 1996, *111*(1), pp. 165–94.

Cooper, Russell and John, Andrew. "Coordinating Coordination Failures in Keynesian Models." *Quarterly Journal of Economics*, August 1988, *103*(3), pp. 441–63.

Crawford, Vincent. "Adaptive Dynamics in Coordination Games." *Econometrica*, January 1995, *63*(1), pp. 103–43.

———. "Theory and Experiment in the Analysis of Strategic Interaction," in David Kreps and Ken Wallis, eds., *Advances in economics and econometrics: theory and applications*, Vol. I, Econometric Society Monographs No. 27. Cambridge: Cambridge University Press, 1997, pp. 206–42.

Crawford, Vincent and Broseta, Bruno. "What Price Coordination? Auctioning the Right to Play as a Form of Preplay Communication." Department of Economics Discussion Paper No. 95-41, University of California, San Diego, November 1995.

Grether, David and Maddala, G. S. "A Time Series Model with Qualitative Variables," in M. Deistler, E. Furst, and G. Schwodiauer, eds., *Games, economic dynamics, and*

series analysis. Vienna: Physica Verlag, 1982, pp. 291–305.

Heckman, James. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica*, July 1974, *42*(4), pp. 679–94.

Kim, Yong-Gwan. "Evolutionary Analyses of Tacit Communication in Van Huyck, Battalio, and Beil's Game Experiments." *Games and Economic Behavior*, October 1996, *16*(2), pp. 218–37.

Maddala, G. S. *Limited-dependent and qualitative variables in econometrics*, Econometric Society Monographs No. 3. Cambridge: Cambridge University Press, 1983.

Manski, Charles and McFadden, Daniel. "Alternative Estimators and Sample Designs for Discrete Choice Analysis," in Charles Manski and Daniel McFadden, eds., *Structural analysis of discrete data with econometric applications*. Cambridge, MA: MIT Press, 1981, pp. 2–50.

McCabe, Kevin; Rassenti, Stephen and Smith, Vernon. "Auction Institutional Design: Theory and Behavior of Simultaneous Multiple-Unit Generalizations of the Dutch and English Auctions." *American Economic Review*, December 1990, *80*(5), pp. 1276–83.

McFadden, Daniel. "Econometric Analysis of Qualitative Response Models," in Zvi Griliches and Michael Intriligator, eds., *Handbook of econometrics*, Vol. II. New York: Elsevier Science Publishing, 1984, pp. 1396–456.

Milgrom, Paul and Weber, Robert. "A Theory of Auctions and Competitive Bidding." *Econometrica*, September 1982, *50*(5), pp. 1089–122.

Nevel'son, M. B. and Has'minskii, R. Z. *Stochastic approximation and recursive estimation*, Vol. 47, Translations of Mathematical Monographs. Providence, RI: American Mathematical Society, 1973.

Pudney, Stephen. *Modelling individual choice: The econometrics of corners, kinks, and holes*. Oxford: Blackwell, 1989.

Roth, Alvin and Erev, Ido. "Learning in Extensive-Form Games: Experimental Data and Simple Dynamic Models in the Intermediate Term." *Games and Economic Behavior*, January 1995, *8*(1), pp. 164–212.

Schelling, Thomas. *The strategy of conflict*, 1st Ed. Cambridge, MA: Harvard University Press, 1960.

Teichroew, D. "Tables of Expected Values of Order Statistics and Products of Order Statistics for Samples of Size Twenty and Less from the Normal Distribution." *Annals of Mathematical Statistics*, June 1956, *27*(2), pp. 410–26.

Van Huyck, John; Battalio, Raymond and Beil, Richard. "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure." *American Economic Review*, March 1990, *80*(1), pp. 234–48.

______. "Strategic Uncertainty, Equilibrium Selection Principles, and Coordination Failure in Average Opinion Games." *Quarterly Journal of Economics*, August 1991, *106*(3), pp. 885–910.

______. "Asset Markets as an Equilibrium Selection Mechanism: Coordination Failure, Game Form Auctions, and Tacit Communication." *Games and Economic Behavior*, July 1993, *5*(3), pp. 485–504.

Weber, Roberto. "The Role of Alternative Market Institutions on Equilibrium Selection in Coordination Games." B.A. honors thesis, Texas A&M University, 1994.

Woodford, Michael. "Learning to Believe in Sunspots." *Econometrica*, March 1990, *58*(2), pp. 277–307.

The International Ramifications of Tax Reforms: Supply-Side Economics in a Global Economy

By ENRIQUE G. MENDOZA AND LINDA L. TESAR*

This paper studies tax reforms in a dynamic model of a global economy calibrated to current U.S. and European tax policies. World capital markets add consumption-smoothing and income-redistribution effects that alter closed-economy predictions. In the absence of taxes on foreign interest, welfare gains of eliminating U.S. income taxes are enlarged by up to 34 percent, at the expense of European losses caused by transitional declines in consumption and leisure, and a permanent capital outflow. In contrast, if foreign interest is taxed, the same tax reform reduces U.S. welfare 0.7 percent and increases European welfare 1.8 percent. (*JEL* H87, H21, H23, F41)

> Up to this point I have taken no account of international relations .... When this assumption is removed, several new and large problems arise ... it may be feasible for a man subjected to taxation in a taxed area to make use of an untaxed area in such a way as to reduce the fiscal burden imposed upon him .... New and so far unexamined dangers are threatened. It is clearly important to gauge, so far as we can, the scope and range of these in the particular case of our own country (A. C. Pigou, 1947 p. 165).

The research program on quantitative assessments of tax reforms initiated by Robert E. Lucas Jr.'s (1990) lecture on supply-side economics concluded that far-reaching tax reforms, designed to eliminate savings and investment distortions, produce large social welfare gains (see, for example, Robert G. King and Sergio T. Rebelo, 1990; Jeremy Greenwood and Gregory W. Huffman, 1991; Thomas F. Cooley and Gary D. Hansen, 1992; V. V. Chari et al., 1994). The welfare gains are large despite the transitional cost incurred in the process of expanding the capital stock from the lower level of a heavily tax-distorted economy to the higher level of a tax-reformed economy. Lucas estimated that replacing the capital income tax with a higher labor income tax would increase consumption per capita by about 1 percent per year, and Cooley and Hansen showed that the increase can exceed [illegible] percent if a consumption tax is used instead. Gains of this magnitude dwarf the benefits of other major policy endeavors—such as output and price stabilization—and, in Lucas's view, constitute the "largest genuinely free lunch" ever provided by quantitative welfare economics.[1]

The academic enthusiasm for revamping the tax system is shared by many policy makers in industrial countries, particularly in France, Germany, Japan, and the United States. In the United States, for instance, Congress created the Kemp Commission on Growth and Tax Reform which published a report in 1996 focusing on three major tax-reform proposals: the "universal savings allowance," which provides deductions for all income added to savings, the "flat tax," which cuts income

* Mendoza: Department of Economics, Duke University, Durham, NC 27708; Tesar: Department of Economics, University of Michigan, Ann Arbor, MI 48109. Comments and suggestions by participants at the 1995 NBER Summer Institute, the World Congress of the Econometric Society, the Canadian Macro Study Group Meetings, and the June 1995 Meeting of the Panel of Economic Advisers of the Congressional Budget Office are gratefully acknowledged.

[1] Recently, S. Rao Aiyagari (1995) showed that eliminating the capital income tax can be suboptimal if there are borrowing constraints and incomplete markets.

taxes to a single 20–25 percent rate and makes up revenue losses with growth gains and higher indirect taxes, and the "consumption-tax-only," which envisages replacing completely the federal income tax with a value-added tax.

Despite the strong interest in radical tax reform on the part of researchers and policy makers, Pigou's concern for analyzing the international ramifications of tax reforms and quantifying their impact has remained largely ignored for over half a century. In fact, all existing quantitative studies in the tradition of Lucas's supply-side lecture assume that international financial markets do not exist. This is sharply at odds with the unprecedented globalization of capital markets that the world has experienced recently and with the continuing process of trade integration. The emergence of large and sophisticated global markets brings to center stage Pigou's concern for understanding how tax-reform assessments vary when the public has access to world trade. This is the central theme of this paper. In particular, the paper undertakes a quantitative examination of the effects of tax reforms from the perspective of a two-country dynamic macroeconomic model with fully integrated capital and goods markets, and calibrated to reflect current tax policies of the United States and Europe.

World financial markets play a key role in both the positive and normative aspects of tax reforms. Without access to external borrowing, households finance the accumulation of capital during the tax-reform transition by sacrificing consumption and leisure, and hence bear the large social costs associated with transitional dynamics. In contrast, world financial markets provide mechanisms for sharing the costs and benefits of tax reforms internationally, and hence produce sizable global spillovers along the lines examined in the analytical work of Jacob A. Frenkel and Assaf Razin (1992) and Greenwood and Kent P. Kimbrough (1985).

Our quantitative analysis shows that trade in world financial markets magnifies the benefits of tax reforms. The net welfare gain of a tax reform replacing the capital income tax with a consumption tax in the United States is [illegible] percent in the open-economy model, nearly 34 percent larger than in a closed-economy setting. A similar reform eliminating the labor income tax produces a welfare gain 10 percent larger for an open economy than for a closed economy. In both cases, the consumption tax increases sharply in order to satisfy the government's intertemporal budget constraint, keeping constant the levels of government expenditures and welfare payments. Most significantly, the transition paths for consumption and work effort change dramatically in the open economy. These calculations assume no change in the long-run growth rate of the economy, so the results do not rely on a permanently higher growth rate.

The international transmission of tax reforms operates through two key effects. The first is a *smoothing effect* reflected in external borrowing by U.S. households to smooth intertemporally the sacrifice of consumption and leisure implicit in the cost of the transition. As a result, the United States runs sizable trade and fiscal deficits in the short run and the cost of transitional dynamics falls sharply, from 7.6 percent in a closed economy to 3.4 percent in an open economy, when the capital income is replaced with a consumption tax. The second effect is a long-run *income-redistribution effect.* This effect captures the notion that the debt accumulated during the transition is serviced by a larger trade surplus in the long run. This mechanism transfers part of the long-run gains of the tax reform abroad. Hence the utility gain measured by comparing pre- and post-tax-reform steady states, ignoring the costs of transitional dynamics, is smaller in an open economy than in a closed economy.

The smoothing and income-redistribution effects produce international spillovers of domestic tax reforms, which affect the dynamics of foreign borrowing and the world interest rate, and hence cause large global externalities in response to unilateral tax-policy choices. In fact, the additional gains accruing to U.S. residents by borrowing internationally are matched by welfare losses in Europe, so tax reforms in an open economy are not "a genuine free lunch," as Lucas (1990) concluded. In the short run, the smoothing effect induces European households to reoptimize their portfolios from physical capital into international bonds, leading to a large capital outflow, and to consume less and work harder in order to

generate the trade surplus to match the U.S. deficit. These movements affect adversely Europe's welfare, and are only partially offset by the long-run income-redistribution effect. The latter leads to a long-run increase in European consumption and leisure, although Europe's capital stock falls permanently. These externalities thus provide incentives for strategic behavior by fiscal authorities similar to those that motivated the large theoretical literature on world tax competition (see Torsten Persson and Guido Tabellini, 1995), but to date no attempts have been made to quantify their magnitude.

We also examine how the international ramifications of tax reforms depend on the general structure of tax policies. Three important lessons emerge from this analysis. First, taxes on foreign interest income alter the incentives for external borrowing by adding a new margin of distortion between the intertemporal marginal rate of substitution in consumption and the world real interest rate. As a result, the distribution of welfare gains of tax reforms across countries can change dramatically. A tax reform that eliminates the U.S. domestic capital income tax, leaving in place a tax on foreign interest income, causes a 0.7-percent welfare loss in the United States and a 1.7-percent welfare gain in Europe. Second, a tax reform in the United States lowers European tax revenue, and hence may force Europe to increase distortionary tax rates in order to maintain intertemporal fiscal balance. In the best-case scenario in which the European policy response is to increase the consumption tax, the European welfare loss induced by the abolition of the U.S. capital income tax is doubled. Third, despite these large global externalities, it is possible to identify simple worldwide tax reforms that produce Pareto improvements. For example, the United States and Europe can obtain sizable welfare gains if they jointly undertake the elimination of the capital income tax in favor of higher consumption taxes.

The paper proceeds as follows. Section I describes the model and discusses the numerical solution method. Section II conducts tax-reform experiments, and explores the implications of alternative tax-policy scenarios and changes in key parameters. Section III concludes.

I. A Dynamic Macroeconomic Model for International Tax-Policy Analysis

A. *Households, Firms, the Public Sector, and Financial Markets*

The analytical framework is a dynamic neoclassical, two-country model. Both countries produce a single composite commodity, and trade both this good and real one-period bonds, issued by the private sector in each country, in perfectly competitive international markets. The description of the model is based on home-country decisions and, when needed, foreign-country decisions are introduced using asterisks to denote foreign variables.

The long-run rate of output growth is exogenous at a rate γ, which is common across countries and across expenditure flows within each country, and is driven by labor-augmenting technological change. This assumption restricts the permissible set of functional forms for preferences and technology to the class of functions that supports balanced growth. The specification of the model is simplified by transforming all variables, except employment and leisure, into stationary variables by dividing them through by the state of technological progress. The transformed variables are written in lower case. This detrending method also requires well-known definitions of the subjective discount factor and laws of motion for asset accumulation. The paper focuses, without loss of generality, on the competitive equilibrium of the detrended model.

Each country is inhabited by identical, infinitely lived individuals. The representative household in the home country maximizes a conventional isoelastic lifetime utility function over intertemporal sequences of consumption (c_t) and leisure (L_t):

$$(1)\quad \sum_{t=0}^{\infty} \beta^t \frac{[c_t L_t^a]^{1-\sigma}}{1-\sigma}, \quad \sigma > 1, \quad a > 0.$$

The stationary transformation of the model requires that β be defined as $\beta \equiv B(1+\gamma)$, where B is the true subjective discount factor.

The household maximizes (1) subject to the following sequence of budget constraints,

ing as given relative prices and fiscal policy variables, for each date $t = 0, \ldots, \infty$:

$$\begin{aligned}(2)\quad & (1+\tau_c)c_t + (1+\gamma) \\ & \quad \times [k_{t+1} + R_t b_{t+1} + R_t^g d_{t+1}] \\ & = (1-\tau_n)w_t N_t \\ & \quad + [1 + (1-\tau_k)(r_t - \delta)]k_t \\ & \quad - \Psi(k_t, x_t) + b_t + d_t.\end{aligned}$$

The left-hand side of (2) represents the household's expenditures, which include purchases of consumption goods, inclusive of a sales tax, τ_c, capital goods, k_{t+1}, private international bonds, b_{t+1}, and domestic government bonds, d_{t+1}. For simplicity, bonds are represented as discounted bonds, so the real returns of private and public bonds are $(1/R_t) - 1$ and $(1/R_t^g) - 1$, respectively. The right-hand side of (2) is the household's disposable factor and nonfactor income. Factor income is derived from supplying capital (k_t) and labor (N_t) to firms at pre-tax rental rates w_t and r_t, and taxed at rates τ_n and τ_k, with a provision for a depreciation allowance. Installed capital depreciates at a fixed rate δ, and additions to installed capital incur capital-adjustment costs $\Psi(k_t, x_t)$, where x_t is net investment $(1+\gamma)k_{t+1} - (1-\delta)k_t$. These costs, or similar frictions like gestation lags, are required in dynamic open-economy models to differentiate physical from financial assets, and thereby prevent the instantaneous adjustment of the domestic marginal product of capital to the world interest rate (see Mendoza, 1991). Without adjustment costs, these models cannot produce transitional dynamics and predict unrealistically large swings in investment rates and current account balances. The last two terms in (2) represent nonfactor income derived from public and private bonds. This income is not taxed, but we show later that allowing for taxation of foreign interest income has important implications. Households also face a no-Ponzi-game restriction, $\mathrm{Lim}_{T\to\infty} (\Pi_{t=0}^{T} R_t) b_T = 0$, which together with (2) implies that the present value of household disposable factor income equals that of expenditures plus any initial bond holdings b_0.

Implicit in equation (2) is the assumption that domestic capital and public debt are owned only by domestic households. This is an extreme assumption, but it has the advantage that it allows the model to support competitive equilibria in which there is free international trade in private bonds and differing capital income tax rates on residents of the two countries. We explain below that this is not possible with cross-border trading on equity or public debt (see also Frenkel et al., 1992).

The household's constraint on the allocation of time between labor and leisure is:

$$(3)\qquad L_t + N_t = 1$$

where the total number of hours is normalized to one. Labor is immobile across countries.

Firms maximize profits subject to constant-returns-to-scale technological constraints, taking as given factor prices. Thus, firms employ inputs according to marginal productivity rules and earn zero profits in equilibrium. The production function is Cobb-Douglas:

$$(4)\qquad F(k_t, N_t) = k_t^{1-\alpha} N_t^{\alpha}.$$

Fiscal policy is represented by an intertemporal sequence of unproductive government expenditures, g_t for $t = 0, \ldots, \infty$, and a set of tax rates. The date-t government budget constraint is:

$$\begin{aligned}(5)\quad g_t + d_t = {} & \tau_k(r_t - \delta)k_t + \tau_n w_t N_t \\ & + \tau_c c_t + (1+\gamma)R_t^g d_{t+1}.\end{aligned}$$

The left-hand side of (5) represents uses of government revenue (i.e., goods purchases and debt payments). The right-hand side includes tax revenue and newly issued debt. Since government purchases and tax rates are exogenous policy choices, the government is assumed to issue new debt as needed to satisfy its budget constraint. Government also faces a no-Ponzi-game constraint, $\mathrm{Lim}_{T\to\infty} (\Pi_{t=0}^{T} R_t^g) d_T = 0$, which jointly with (5) implies that the present value of government expenditures equals the present value of tax revenue plus the initial stock of public debt d_0. Moreover, (2), (5), and

the no-Ponzi-game constraints imply that the present value of the trade balance equals b_0.

Public debt in this model is "Ricardian" in the sense that, given d_0 and policy choices on government purchases and tax rates, the competitive equilibrium can be represented with the government debt path dictated by (5), or with adjustments in lump-sum transfers to households, tr_t, by the amount required to balance the government budget constraint each period:

$$(6) \quad tr_t = \tau_k(r_t - \delta)k_t + \tau_n w_t N_t + \tau_c c_t - g_t.$$

The equivalence between the intertemporal sequences of debt and transfers is clear taking as given d_0 and noting that (5) and (6) imply $tr_t = d_t - (1+\gamma)R_t^g d_{t+1}$. This framework also allows for a constant level of exogenous government transfers to households, representing subsidies and welfare programs, which can be denoted as T and added as an extra right-hand-side term in (2), (5), and (6).

The market-clearing conditions for the world markets of goods and bonds are:

$$(7) \quad F(k_t, N_t) + F(k_t^*, N_t^*) = c_t + c_t^* + x_t + x_t^* + \Psi(k_t, x_t) + \Psi(k_t^*, x_t^*) + g_t + g_t^*,$$

$$(8) \quad b_t + b_t^* = 0.$$

The model's competitive equilibrium is given by sequences of prices $[r_t, r_t^*, R_t, w_t, w_t^*]_{t=0}^{\infty}$ and allocations $[k_{t+1}, k_{t+1}^*, b_{t+1}, b_{t+1}^*, N_t, N_t^*, c_t, c_t^*, L_t, L_t^*, tr_t, tr_t^*]_{t=0}^{\infty}$ that satisfy the first-order conditions of the optimization problems faced by households and firms, the constraints of households and governments, and conditions (7)–(8)—given $[k_0, k_0^*, b_0, b_0^*, d_0, d_0^*]$ and the choice of fiscal instruments.

B. *The International Transmission of Tax Reforms*

The first-order conditions that characterize optimal decisions provide important intuition for understanding the international ramifications of tax reforms. The first-order conditions for investment and foreign bonds in each country, ignoring adjustment costs, are:

$$(9) \quad \frac{(1+\gamma)U_1(c_t, L_t)}{\beta U_1(c_{t+1}, L_{t+1})} = (1-\tau_k)(F_1(k_{t+1}, N_{t+1}) - \delta) + 1 = R_t^{-1},$$

$$(10) \quad \frac{(1+\gamma)U_1(c_t^*, L_t^*)}{\beta U_1(c_{t+1}^*, L_{t+1}^*)} = (1-\tau_k^*)(F_1(k_{t+1}^*, N_{t+1}^*) - \delta) + 1 = R_t^{-1}.$$

The model supports equilibria where countries trade private bonds and $\tau_k \neq \tau_k^*$. This follows from two implications of conditions (9)–(10). First, trade in bonds implies that countries face a common intertemporal relative price of consumption R_t^{-1}, and hence growth-adjusted intertemporal marginal rates of substitution in consumption are equalized. Second, the optimal portfolio allocation across capital and private bonds requires that the post-tax net marginal products of capital are also equalized across countries. As a result, differences in capital income taxes are offset by differences in pre-tax net marginal products of capital. This cannot occur if countries trade equity and tax capital income according to the residence principle (i.e., home households pay τ_k on their holdings of k and k^*). In this case, both pre-tax and post-tax returns on capital are equalized, and hence a world competitive equilibrium requires $\tau_k = \tau_k^*$. A similar result applies to world trade in public debt: equilibria with $\tau_k \neq \tau_k^*$ can be supported only if there is no world trade in public debt, as we assumed, or if public debt is internationally traded but interest payments on it are tax free. In the latter case, however, b and d would be perfect substitutes and there would not be well-defined portfolio shares assigned to each debt instrument. The assumption that neither equity nor public debt are traded across countries is restrictive, but its has two advantages: it keeps the model simple and it prevents the model from being clearly at odds with the

differences in domestic capital income tax rates across countries that we document below.

To understand the international transmission of tax reforms, consider next the implications of a permanent, unanticipated cut in the home-country capital income tax. Conditions (9)–(10) imply that, for a given world interest rate $R_t^{-1} - 1$, the tax cut increases the net-of-tax domestic marginal product of capital, and hence arbitrage with the bond market implies that k_{t+1} must rise to restore equilibrium. There is no direct "arbitrage" effect on foreign capital, since equity is not traded globally, but to the extent that domestic households borrow from abroad to finance the increase in k_{t+1} they induce a capital outflow from the foreign country. Moreover, since we are dealing with two large countries, when one country changes its net foreign debt it also alters the world interest rate. Interest rate changes are temporary, however, because the long-run interest rate r is pinned down by the steady-state condition $r = \rho - \gamma\sigma$, where ρ is the rate of time preference, defined as $\rho \equiv \beta^{-1} - 1$.

The above intuition implies that transitional and long-run changes in international borrowing, and transitional changes in the interest rate, are the channels for global transmission of tax reforms in the model. We condense these channels into a *smoothing effect* and an *income-redistribution effect*. The first effect refers to the resources that a county obtains from world markets to lessen the cost of the transitional dynamics in the initial stages of a tax reform, and can be measured by transitional changes in net exports—which combine both interest rate changes and changes in net foreign asset positions. The second effect reflects the redistribution of income across countries that occurs because a country that accumulates foreign debt during the transitional dynamics maintains a long-run trade surplus to service that debt. Since the long-run interest rate is invariant to tax changes, this second effect captures only long-run changes in foreign asset positions.

Note that an implication of the above-mentioned effects is that the benefits that a tax-cutting country extracts from world capital markets depend on how its borrowing decisions affect R_t. In the extreme case of a small open economy, without adjustment costs, R is constant and a cut in τ_k is matched by an immediate and large increase in x_t entirely financed by external borrowing. A small country will borrow more, and at a lower cost, than a large country because for the small country the world supply of capital is infinitely elastic.

Labor income taxes and consumption taxes also have international implications, but they are less direct. Changes in these tax rates operate first through the consumption-leisure trade-off:

$$(11) \qquad \frac{U_2(c_t, L_t)}{U_1(c_t, L_t)} = \frac{(1 - \tau_n)}{(1 + \tau_c)} w_t.$$

Given Cobb-Douglas production technologies, the resulting distortions on labor supply affect the marginal products of both capital and labor, and the effect on the former triggers the international transmission mechanisms described earlier. Note also that τ_n and τ_c jointly distort the marginal rate of substitution between consumption and leisure, but their impact on tax revenue, household income, and welfare differs, as shown below.

C. *Income Tax Reforms: Calibration and Solution Method*

We study tax reforms in which the government undertakes a permanent, unanticipated reduction in time-invariant factor income taxes at $t = 0$.[2] Government expenditures and exogenous welfare transfers remain fixed at the levels g_0 and T, respectively. The revenue lost due to income tax cuts is replaced by increasing τ_c so that the present value of tax revenue equals that of government expenses (assuming, without loss of generality, that $d_0 = 0$). The government adjusts tr_t (or issues debt) along the transition path as needed to make up

[2] We follow Lucas (1990) and Cooley and Hansen (1992) in limiting the analysis to changes in time-invariant tax rates. Cooley and Hansen provide evidence suggesting that the extra gains made with time-variant taxes relative to time-invariant taxes can be small.

for any shortfall or excess of tax revenue over expenses. Following Lucas (1987, 1990), the net welfare effect of the reforms is measured as the constant percentage increase in c_t, for $t = 0, \ldots, \infty$, that leaves households indifferent between the lifetime utility obtained by remaining in the pre-reform equilibrium, and the lifetime utility obtained by undertaking the tax reform, inclusive of the transitional dynamics of c_t and L_t. This net gain is also decomposed into a long-run gain, measured by comparing lifetime utility across pre- and post-tax-reform steady states, and the short-run cost of the transitional dynamics.

Numerical solutions of the tax-reform experiments involve the computation of: (a) long-run, balanced-growth equilibria before and after the tax reform, and (b) transitional dynamics between pre- and post-tax-reform steady states. The computation of the pre-reform equilibrium is based on a calibration exercise similar to those undertaken in closed-economy studies. In contrast, the computation of the post-tax-reform, balanced-growth equilibrium and the transitional dynamics differs markedly from closed-economy studies because in the open economy the two must be solved simultaneously. This is because, while closed-economy models feature explicit steady-state solutions invariant to initial conditions, in open-economy models there are no explicit steady-state solutions for the allocations of consumption and private bonds across countries, and the dynamics and post-tax-reform steady state of b vary with initial conditions. David Lipton et al. (1982) and Lipton and Jeffrey Sachs (1983) examined similar cases in which steady-state foreign asset positions are part of a two-point boundary problem in the context of dynamic, open-economy IS-LM models.

Calibration of the Pre-tax-Reform Equilibrium.—The following conditions summarize the long-run equilibrium of the home country along the balanced-growth path:

$$\frac{k}{y} = \frac{\beta(1-\alpha)(1-\tau_k)}{(1+\gamma) - \beta[1 - \delta(1-\tau_k)]}, \tag{12}$$

$$\frac{x}{y} = (\gamma + \delta)\frac{k}{y}, \tag{13}$$

$$\frac{c}{y} = 1 - \frac{x}{y} - \frac{g}{y} - \frac{tb}{y}, \tag{14}$$

$$N = \frac{\dfrac{1-\tau_n}{1+\tau_c}\alpha}{a\dfrac{c}{y} + \dfrac{1-\tau_n}{1+\tau_c}\alpha}. \tag{15}$$

Condition (12) is the steady-state version of the Euler equation for capital, and expresses the capital-GDP ratio, k/y, as a function of preference and technology parameters and τ_k. Equation (13) is the law of motion for capital accumulation, and determines the steady-state investment rate, x/y, as a function of γ, δ, and k/y. Condition (14) uses the budget constraint to define the consumption-output ratio, c/y, as a function of x/y and the GDP shares of government purchases and net exports (g/y and tb/y, respectively). Since along the balanced-growth path $tb/y = (\beta - 1)b/y$, the private debt ratio b/y is a simple transformation of tb/y. Condition (15) follows from (11) and sets N as a function of c/y; τ_n, τ_c, α, and a. In preparation for the analysis of Section II, note that (15) also determines N_t at any date in the equilibrium path, with c/y replaced by c_t/y_t. Note also that τ_k affects both x/y and, through its effect on c/y, the supply of labor, while τ_n and τ_c do not affect x/y.

Formally, equations (12)–(15) are an underidentified system of four equations with five unknowns (k/y, x/y, c/y, tb/y, and N). Equations (12)–(13) are block recursive and determine k/y and x/y exactly as in a closed-economy model, but (14)–(15) cannot determine c/y, tb/y, and N. Thus, balanced-growth equilibria of these variables are not pinned down by steady-state conditions. The calibration of the pre-tax-reform equilibrium circumvents this problem by taking tb/y from the data. More precisely, the system (12)–(15) is solved for δ, β, a, and c/y, given the values of other preference and technology parameters, tax rates, and long-run averages of k/y, tb/y, g/y, and N taken from actual data. The home country is calibrated to U.S. data, and the foreign country corresponds to European aggregates measured as arithmetic averages of data for France, Germany, Italy, and

TABLE 1—PARAMETER VALUES AND PRE-TAX-REFORM, STEADY-STATE ALLOCATIONS

Parameter values:

Technology and preferences:

δ	α	γ	η	B	σ	a
0.0161	0.64	0.0039	10	0.993	2	2.675

Fiscal policy parameters (in percent)[a]

	United States	Europe
τ_k	41.5	34.3
τ_n	29.1	38.2
τ_c	4.4	15.8
g/y[b]	19.0	21.0

Pre-tax-reform, balanced-growth allocations (GDP ratios):

	United States		Europe	
	Data	Model	Data	Model
c/y	0.65	0.65	0.60	0.59
x/y[c]	0.17	0.17	0.17	0.18
tb/y	−0.01	−0.01	0.01	0.01
Tax revenue	0.28	0.30	0.36	0.41
Net transfers[d]	0.14	0.12	0.24	0.20

Notes: Figures in "Data" columns are averages for the 1968–1991 period, based on national accounts and tax revenue data from *OECD National Accounts* and *Revenue Statistics.*

[a] Tax rates are 1991 estimates computed as in Mendoza et al. (1994).

[b] Government expenditures (including public investment) at the general government level.

[c] Private investment rate. Data not available for Italy. The figure shown for Europe is the average of the private investment-GDP ratio in France, Germany, and the United Kingdom.

[d] Subsidies and all current transfers. Data for Italy and the UK are for the 1980–1988 period.

United Kingdom. Long-run GDP ratios are based on data from the OECD's *National Accounts* and *Revenue Statistics* and on estimates reported by Cooley and Hansen (1992).

The parameter values used to calibrate the model at a quarterly frequency are shown in Table 1. The values of g/y and g/y^* are easily derived from the data, and are estimated at 19 percent for the United States and 21 percent for Europe, including public investment. In contrast, obtaining macroeconomic estimates of tax rates is difficult due to complex international differences in tax codes (credits, exemptions, deductions, etc.) and to the progressivity and nonlinearity of tax schedules. In previous work with Assaf Razin (Mendoza et al., 1994), we estimated tax rates for Europe and the United States over the 1968–1990 period by combining detailed tax revenue statistics with information from the aggregate balance sheets of households, corporations, and government from national income accounts. Figure 1 plots the estimated tax rates. We set the pre-tax-reform tax rates equal to 1990 values. The estimates suggest that τ_k is larger in the United States than in Europe (41.5 percent compared to 34.3 percent) and, conversely, τ_n and τ_c are larger in Europe than in the United States (15.8 versus 4.4 percent for the consumption tax and 38.2 versus 29.1 percent for the labor income tax). These tax rates capture the widely accepted view that,

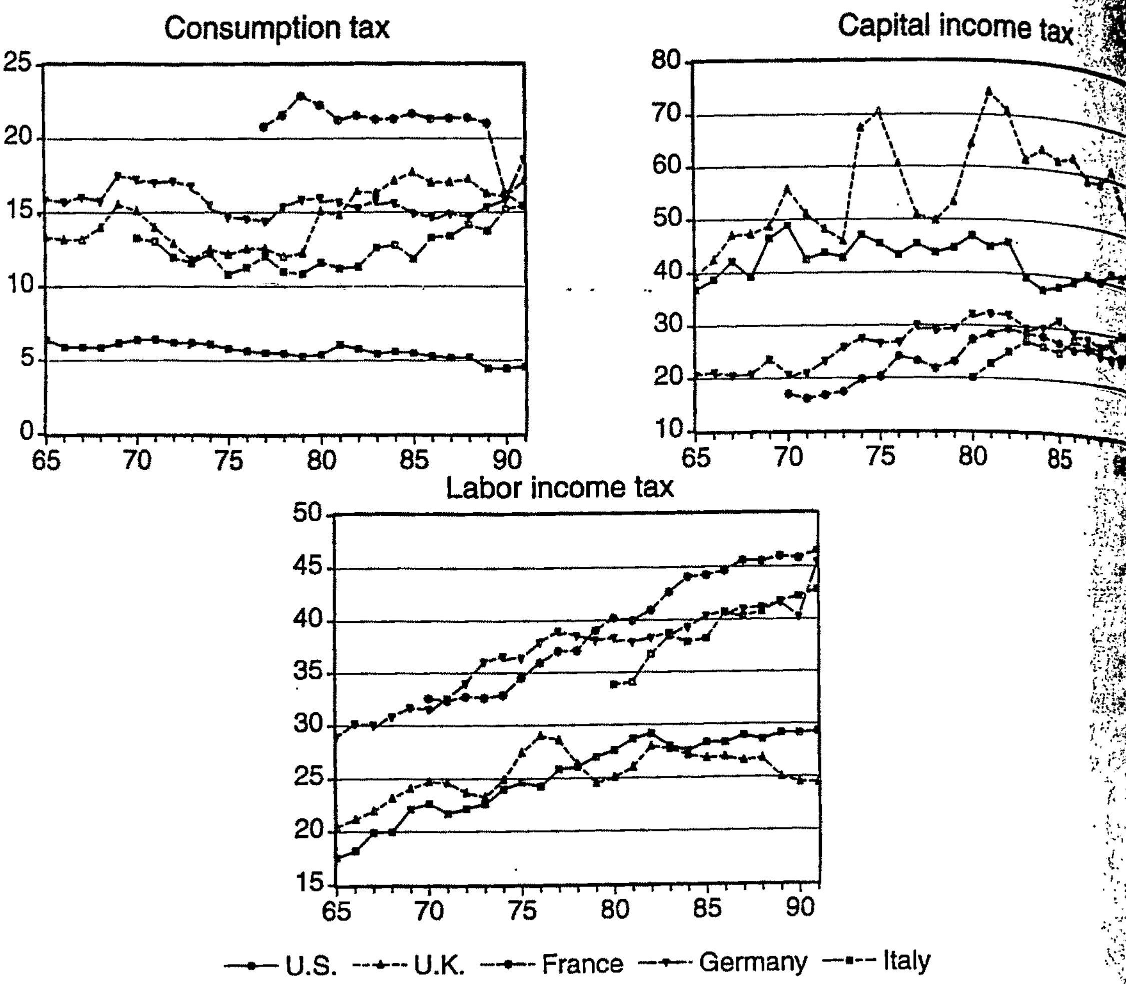


FIGURE 1. EFFECTIVE TAX RATES IN EUROPE AND THE UNITED STATES

compared to Europe, the United States taxes consumption and labor income less heavily than capital income.

The rest of the calibration follows standard practice in real-business-cycle analysis (see Chs. 1 and 7 of Cooley, 1995). The per capita GDP growth rate is set to $\gamma = 1.56$ percent per annum (0.39 percent per quarter) and the intertemporal elasticity of substitution is set at $^1/_2$ (i.e., $\sigma = 2$). We measured the annual x/y ratio, adjusted to exclude public investment, at 0.17, which is similar to the estimate Cooley and Hansen (1992) obtained for the post-war ratio of fixed nonresidential investment relative to corporate GDP. The quarterly k/y ratio is set to 2.16 (8.62 annually), which is also in line with the figure Cooley and Hansen used (2.13). We also followed these authors in setting the labor share, α, equal to 0.64. Given these ratios and parameter values and the tax rates, conditions (14) and (15) imply values of δ equal to 1.61 and B equal to 0.99 (where $\beta \equiv B(1 + \gamma)^{1-\sigma}$). The implied value of r is 6.1 percent per annum. Given x/y and g/y, conditions (14) and (15) jointly determine a value of a consistent with a labor allocation equal to 20 percent of time, and a ratio c/y consistent with $tb/y = -1$ percent, both conforming to U.S. data. This implies $a = 2.675$ and $c/y = 0.65$. Preference and technology parameters are set identical across the United States and Europe to highlight the effects of asymmetries in fiscal policy. Thus, we adopt the same values of a, α, β, δ, σ, and γ for Europe and allow k/y^*, x/y^*, and N^* to adjust accordingly.

Table 1 also reports the measures of c/y, x/y, tb/y, and the GDP ratios of tax revenue and net government transfers for both countries in the data and in the model. The calibration was set to mimic the U.S. ratios g/y, x/y, and tb/y, but all other ratios are endogenous solutions of the steady-state system of the model. Hence, the fact that the model's ratios are roughly in line with the data suggests that the pre-tax-reform equilibrium is a reasonable platform for tax-reform analysis. Note that in the data tax revenue exceeds government purchases by over 10 percentage points of GDP, in part because net government transfers to households amount to 14 and 24 percent of GDP in the United States and Europe, respectively. These transfers measure subsidies and payments on account of large welfare, health-care, and other entitlement programs. As explained, the model captures this fact by allowing for an exogenous rebate of a part of tax revenue that is kept constant at the pre-tax-reform level in the tax-reform experiments. Consequently, the transitional fluctuations in net transfers computed below correspond exactly to adjustments in the stock of public debt.

The model also features capital adjustment costs of the following convex, quadratic form:

$$\Psi(k_t, x_t) = \frac{\eta}{2}\left(\frac{x_t}{k_t} - z\right)^2 k_t \tag{16}$$

where $z \equiv x/k$, so that adjustment costs are zero in steady state. The value of η is set so that the average convergence rate of GDP to the long-run, balanced-growth path corresponds to empirical estimates that set conditional growth-convergence coefficients at about 2–3 percent (see Robert J. Barro and Xavier Sala-i-Martin, 1995).[3] This implies $\eta = 10$. Barro and Sala-i-Martin (1995) noted that this approach can yield values of η that exceed empirical evidence from investment equations by a wide margin. However, adjustment costs can also be interpreted as a proxy for other frictions absent from the model—such as the existence of a nontraded goods sector—which slow down aggregate capital accumulation. Mendoza and Tesar (1995) show that a two-country model with nontraded goods, calibrated setting η to the value used in the study of international business cycles by Mendoza (1991), produces similar results for tax-reform assessments as the present model. We focused on the one-sector model here for expositional ease and examined the sensitivity of the numerical simulations to changes in the value of η.

Transitional Dynamics.—The solution of the transitional dynamics of a tax reform requires the simultaneous solution of the paths of foreign debt accumulation and the net foreign asset positions in the post-reform steady state. We propose a method that ensures that post-reform, steady-state bond positions are consistent with debt-accumulation dynamics, starting from initial conditions corresponding to the pre-reform equilibrium. The method blends the King-Plosser-Rebelo (KPR) linear approximation algorithm with an iterative "shooting" routine on foreign debt.[4] We take an initial guess of the long-run bond positions to which countries converge after the tax reform (typically the pre-tax-reform positions), solve (12)–(15) for k/y, x/y, c/y, and N, and linearize around the implied balanced-growth allocations. Then we simulate the transitional dynamics for 2,500 periods using KPR and setting as initial conditions the pre-reform values of the state variables k_0, k_0^*, and b_0. The simulations produce a path of foreign debt dynamics that converges to some long-run bond positions. If these differ from the initial guess, the new results are adopted as a new guess and the process is repeated. The process converges rapidly, in four or five iterations, but the difference between the initial guess and the final outcome can be quite large. For instance, application of the KPR method assuming that the pre- and post-reform bond positions are identical produces welfare gains for the tax reforms examined in Section II about $^1/_3$ smaller

[3] The average convergence rate is $1/T \sum_{t=0}^{T} \ln(y_{t+1}/y)/T$, where y_t reaches the steady state at T.

[4] An alternative to the KPR method is to solve the sequence of Euler equations as in Cooley and Hansen (1992). This is computationally expensive in models with a large number of state variables. Our algorithm can mimic closely the results of the Cooley-Hansen tax-reform experiments.

than the true values. The income-redistribution effect is also underestimated by a large margin—the correct long-run trade balance is a *surplus* of 3.8 percent of GDP, compared to a *deficit* of 1 percent in the incorrect solution.

The solution method requires a second "shooting" routine to ensure that the intertemporal government budget constraint is satisfied. The algorithm checks whether the present values of government expenses and tax revenue are equal. If this equality fails, τ_c is adjusted according to a rule that updates the tax rate as needed to balance the latest estimate of the government's intertemporal budget constraint. Note that, in the long run, the adjustment in τ_c is smaller the larger the rise in c and the larger the rise in y relative to c (i.e., the larger the raise in the base of the consumption tax and the larger the "supply-side" effect enhancing income tax revenue collection). In the short run, however, tax revenue declines with respect to the pre-tax-reform level, and this requires a larger increase in τ_c the larger the transitional declines in consumption and factor incomes. The algorithm starts again with the shooting routine on bonds using the updated τ_c, and the process is repeated until it converges to consistent solutions for both τ_c and long-run bond positions. We assume for now that the foreign country uses lump-sum taxes to offset any global fiscal effects of the home-country tax reform. Later we explore the implications of assuming τ_c^* is adjusted to maintain intertemporal fiscal balance abroad.

II. The International Effects of U.S. Income Tax Reforms

A. *Elimination of U.S. Factor Income Taxes*

The quantitative exploration begins with the analysis of a unilateral reduction in U.S. factor income taxes for both the two-country case and the case in which the U.S. economy is modelled as a closed economy. Since tax-reform proposals differ markedly on the allocation of tax cuts across labor and capital, we consider two basic tax reforms: a reform that replaces the capital income tax with the consumption tax, keeping the labor income tax constant, and a reform that replaces the labor income tax with the consumption tax, leaving constant the capital income tax. This strategy is also intended to illustrate how the international implications of two major income tax reforms differ. As we show, the global spillovers are most relevant in the case of capital income tax reforms, and hence much of the subsequent analysis will focus on this case.

The top part of Table 2 shows the effects of the elimination of the capital income tax. The reform reduces τ_k to zero from its current value in the United States (41.5 percent). Consider first the case in which the United States is a closed economy. The model predicts that τ_c needs to rise to 16.3 percent to maintain fiscal balance. Welfare is 9.8 percent higher in the post-reform steady state than in the pre-reform steady state, but the transition to the new steady state implies a hefty social cost of 7.6 percent. Still, the net welfare gain of the reform is a sizable 2.2 percent.[5] The impact effects show that the tax reform causes c_0 to fall by 8.3 percent and L_0 to rise slightly by 0.1 percentage points. The rise in the consumption tax increases the relative price of labor relative to leisure, causing a substitution effect favoring a fall in labor supply. On the other hand, the adverse income effect reflected in the falling share of consumption relative to output at date 0 favors an increase in labor supply. The two effects almost neutralize each other [see (15)] and result in the small rise in L_0. The fall in the consumption-output ratio is due to the increase of 7.3 percentage points in x_0/y_0, as the process of capital accumulation of the transition to the new long-run equilibrium begins. This process causes a temporary increase in r_0 of 3 basis points in annual terms.

Consider now the results of the open-economy model. The effects of the tax reform are radically different in the following four key dimensions.

(1) *Intertemporal smoothing.*—The cost of the transition is sharply reduced by intertemporal smoothing through borrowing on international capital markets. On impact, tb_0/y_0 falls

[5] This estimate is comparable with the 2.8-percent gain obtained by Cooley and Hansen (1992) for a similar experiment replacing τ_k with τ_c under a slightly different parameterization.

TABLE 2—MACROECONOMIC EFFECTS OF INCOME TAX REFORMS IN THE UNITED STATES

	Closed economy		Open economy			
	United States		United States		Europe	
Variable	Impact effect	Long-run effect	Impact effect	Long-run effect	Impact effect	Long-run effect
	Replacing the U.S. capital income tax with a consumption tax					
New tax rates						
τ_K	0.000	0.000	0.000	0.000		
τ_C	16.333	16.333	14.775	14.775		
Welfare effects						
Transitional cost		−7.644		−3.415		−4.860
+ steady-state gain		9.805		6.309		4.036
= net change		2.161		2.894		−0.824
Percentage changes						
y	−0.456	13.490	−2.795	19.326	4.262	−5.037
c	−8.343	6.991	−3.858	7.099	−4.031	1.082
k	0.000	52.371	0.000	60.206	0.000	−5.037
Percentage point changes						
tb/y	—	—	−9.886	3.817	9.969	−4.944
x/y	7.260	5.899	11.205	5.899	−3.547	0.000
r	0.030	0.000	0.016	0.000	0.016	0.000
L	0.144	0.778	0.867	−0.221	−1.192	0.891
	Replacing the U.S. labor income tax with a consumption tax					
New tax rates						
τ_L	0.000	0.000	0.000	0.000		
τ_C	29.375	29.375	28.795	28.795		
Welfare effects						
Transitional cost		−1.697		−0.696		−1.042
+ steady-state gain		4.835		4.148		0.791
= net change		3.138		3.452		−0.251
Percentage changes						
y	5.921	8.760	5.390	10.045	1.031	−0.998
c	7.684	11.335	8.811	11.500	−0.936	0.215
k	0.000	8.760	0.000	10.045	0.000	−0.998
Percentage point changes						
tb/y	—	—	−1.935	0.878	2.247	−0.940
x/y	0.079	0.000	0.887	0.000	−0.880	0.000
r	0.006	0.000	0.004	0.000	0.004	0.000
L	−1.904	−1.774	−1.713	−2.013	−0.286	0.177

Notes: y = gross domestic produce (GDP), c = consumption, k = capital stock, tb/y = trade balance-GDP ratio, x/y = investment-GDP ratio, r = real interest rate, and L = leisure hours. Figures are changes relative to pre-tax-reform, balanced-growth allocations at quarterly frequencies, except the interest rate, which is annualized. Initial levels in closed and open economies differ slightly because the latter is calibrated to reflect a trade deficit of 1 percent of GDP. Adjusting for this difference has negligible effects.

by almost 10 percentage points to finance the increase in investment and smooth consumption. The fall in c_0 is about $^1/_2$ the size of the decline in the closed-economy case, and r_0 rises by only half as much. The cost of the transitional dynamics falls from 7.6 percent to 3.4 percent.

(2) *Income-redistribution effect.*—The income-redistribution effect results in a smaller steady-state gain in the open economy because the United States runs a permanent trade surplus of 3.8 percent of GDP to service its foreign debt. Still, the welfare gain net of transitional dynamics reaches 2.9 percent, exceeding the 2.2 percent of the closed economy by about $^1/_3$.

(3) *Dynamics of labor supply.*—The impact and long-run effects on labor and leisure differ markedly from the closed-economy case. In the closed economy, L_0 remains nearly constant. In contrast, international borrowing allows L_0 to rise nearly 1 percentage point in the open economy. In the long run, leisure in the closed economy rises by 0.8 percentage points, while in the open economy leisure falls about $^1/_4$ of a percentage point. The extra output the open economy needs in order to maintain the trade surplus that services its enlarged external debt implies that labor supply must be higher, and hence leisure smaller, when world markets are considered [see equations (14) and (15)].

(4) *Adjustment in the consumption tax.*—The higher long-run levels of labor supply and the capital stock in the open economy imply higher levels of GDP and consumption, which yield higher labor income and consumption tax revenues allowing the government to maintain budget balance with a smaller increase in τ_c. This occurs despite larger transitional declines in tax revenue resulting from the short-run falls in labor supply and consumption. The result is that τ_c increases to 14.8 percent in the open economy, about 1.5 percentage points less than in the closed economy.

Figure 2 plots the transitional dynamics induced by the tax reform. The series plotted are percent deviations from the pre-reform steady state for the United States, as a closed economy (dashed line), as an open economy (solid line), and for Europe (dotted line). The model produces simulations for 2,500 quarters, but the plots show only the first 200 quarters. The dynamics of consumption, leisure, utility, the capital stock, GDP, and net exports reflect the intuition developed above, and illustrate the smooth dynamics typical of neoclassical models with isoelastic preferences and technologies. In addition, the figure plots the dynamics of tax revenue and the fiscal balance. Tax revenue falls more and the fiscal deficit widens more during the transition of the open economy than in the closed economy. Interestingly, these results show that the capital income tax reform produces large external and fiscal deficits that persist for over a decade. Thus, tax reform can produce the "twin deficits" phenomenon, and rationalize large fiscal and trade deficits as natural equilibrium outcomes of economic reform.

The open-economy model also predicts that the U.S. capital income tax reform has major consequences abroad. These are observed in the transitional dynamic paths for Europe and in the corresponding columns of Table 2. The additional welfare gain enjoyed by U.S. residents due to international borrowing is matched by a welfare loss in Europe. European consumption and leisure fall on impact by 4 percent and 1.2 percentage points, respectively, and Europe also suffers a permanent decline in its capital stock (of about 5 percent in the long run) resulting from the outflow of financial capital into the United States. Although European households cannot invest directly in the U.S. capital stock, they help finance the accumulation of U.S. capital by shifting the composition of their savings from investing in their own capital to international bonds. Thus, the reduction in the U.S. capital income tax leads to a global reallocation of capital via the bond market. The smoothing and income-redistribution effects are the causes of these externalities, the magnitude of which can be measured by the impact and long-run effects of the reform on the U.S. trade balance and the world interest rate reviewed earlier.

Consider next the implications of the labor income tax reform, summarized in the bottom panel of Table 2. The U.S. labor income tax falls from 29.1 percent to 0 and the consumption tax increases from 4.4 percent to around 29 percent in both closed and open economies

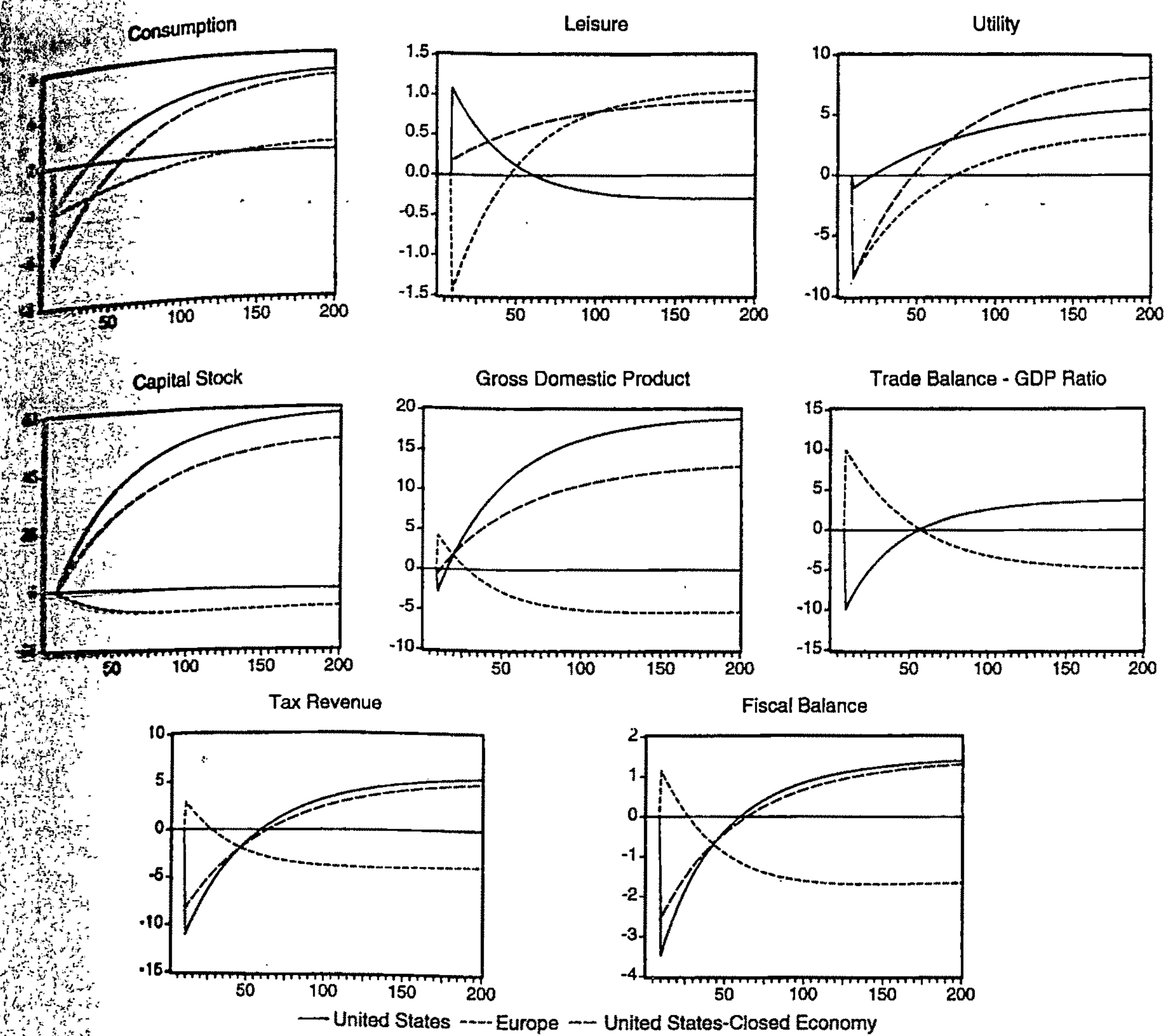


FIGURE 2. TRANSITIONAL DYNAMICS OF A CAPITAL INCOME TAX REFORM IN THE UNITED STATES

Note: Tax revenue and fiscal balance are in percent of GDP. Data plotted are deviations in percent of pre-tax-reform equilibrium, except trade balance-GDP ratio, tax revenue, and fiscal balance, which are percentage-point deviations relative to pre-tax-reform equilibrium.

The results differ from those of the capital income tax reform in three important respects. First, since neither τ_c or τ_n affect the long-run investment rate and the intertemporal-consumption decision margin, this tax reform has no effect on the investment-output ratio in the long run, and has only small effects during the transition. Accordingly, the smoothing and income-redistribution effects triggered by this reform are significantly weaker, as can be seen in the much smaller effects on the trade balance-output ratio and the interest rate—although qualitatively the consequences of international borrowing are the same as before. Second, because both τ_n and τ_c affect the same decision margin (the leisure-consumption choice), and both labor income and consumption are equivalent to about two-thirds of national income, the elimination of τ_n requires a much larger increase in τ_c than the elimination of τ_k. Third, the labor tax is more distortionary than the capital tax, in the sense that eliminating the former produces larger welfare gains in both closed and open economies. This occurs because, although the steady-state welfare gain of the capital tax reform is about two times larger than that of the labor tax reform, the cost of transitional dynamics is nearly five

TABLE 3—MACROECONOMIC EFFECTS OF ELIMINATING THE U.S. CAPITAL INCOME TAX: ALTERNATIVE SCENARIOS

	Home country					Foreign country				
		Impact effects		Long-run effects			Impact effects		Long-run effects	
Model economy	Welfare change	tb/y	c	tb/y	c	Welfare change	tb/y	c	tb/y	c
Benchmark case	2.894	−9.886	−3.858	3.817	7.099	−0.824	9.969	−4.031	−4.944	[illegible]
Alternative tax policies										
World tax reform ($\tau_{\iota} = 15.9$, $\tau_{\iota}^{*} = 25.7$)	2.441	−1.551	−7.361	0.809	7.114	0.830	1.645	−7.824	−0.871	[illegible]
Adjusting τ_{ι}^{*} ($\tau_{\iota}^{*} = 18.4$)	2.921	−10.224	−3.706	3.953	7.093	−1.695	10.385	−5.458	−5.206	[illegible]
Tax on bonds ($\tau_{b} = 1.05$)[a]	−0.676	−8.126	−7.108	6.150	3.046	1.680	8.016	−2.686	−7.865	[illegible]
Sensitivity analysis										
Costless investment	3.760	−40.412	−3.977	2.472	8.517	−1.822	40.570	−6.036	−3.045	[illegible]
Inelastic labor supply	3.271	−8.237	−2.153	2.427	8.143	−0.370	7.920	−5.602	−2.580	[illegible]
Zero trade deficit[b]	2.743	−9.959	−4.057	3.756	6.920	−0.702	10.404	−4.020	−5.299	[illegible]

Notes: Impact effects correspond to changes in the date of the tax reform and long-run effects are changes in post-reform, balanced-growth allocations, both relative to the pre-tax-reform, balanced-growth equilibrium. c is the percentage change in consumption and tb/y is the change in the net exports-GDP ratio measured in percentage points.

[a] This scenario assumes that prior to the tax reform the home country applies the "residence" principle, so that τ_b is set to equate the tax rates on domestic capital income and foreign interest income. The tax reform eliminates τ_k but retains τ_b at the same level of 1.05 percent.

[b] This scenario assumes zero trade deficits in the pre-reform steady state and is calibrated by setting the initial bond positions to 0.

times larger. Consumption falls sharply during the transition of the capital tax reform while it increases sharply in the case of the labor tax reform. The opposite movements are observed in leisure, but given the assumed substitutability of consumption and leisure and the wage elasticity of the latter, households prefer the outcome with the sharp initial consumption boom associated with the labor tax reform.

B. *Key Determinants of the Global Spillovers of Tax Reforms*

The previous analysis showed that a unilateral reduction in factor income taxes by one country has significant implications worldwide. This result was established in an environment where the United States replaced income taxes with consumption taxes, holding constant other key parameters of domestic and foreign tax policy. Thus, the next critical step is to study how the outcome of the analysis varies when these considerations are introduced.

The top panel of Table 3 reports results for three alternative tax-policy scenarios, all for the case of the U.S. capital income tax reform. The benchmark results of Table 2 are also provided to facilitate comparisons. The first alternative tax-policy setting considers the fact that, since the U.S. tax reform reduces European welfare, the tax reform is likely to be "unsustainable," in the sense that it would trigger a response by Europe's tax authorities and would likely lead to a game of global tax competition. Examining this issue in detail

beyond the scope of this paper, but we nevertheless show that it is possible to design a simple worldwide tax reform that constitutes a Pareto improvement over current tax systems.[8] This is done in the second row of Table 3, which assumes that both the United States and Europe replace capital income taxes with higher consumption taxes. Europe makes a 0.[illegible]-percent welfare gain, instead of a loss, and the United States still gains 2.4 percent. The impact and long-run effects of this reform on the trade balance-GDP ratio are much weaker than in the benchmark case because both countries are undergoing the transition to a larger steady-state capital stock. Hence, both countries have incentives to borrow internationally and world financial markets do less to mitigate the costs of the transition. This also implies that the interest rate must rise more in order to clear world markets. Note, however, that this "coordinated tax reform" does not represent a solution to either cooperative or noncooperative tax competition—except for a cooperative outcome based on an additive world welfare function parameterized arbitrarily with the appropriate welfare weights.

The second major change in the tax-policy setting that we examine involves relaxing the assumption that Europe has access to lump-sum taxation as a means to offset the impact of the U.S. tax reform on European tax revenue (which is not trivial, as shown in Figure 3). The third row of Table 3 thus deals with the case in which Europe adjusts its consumption tax to maintain the present value of its tax revenue identical to the unchanged present value of its government expenses. Clearly, Europe must be worse off because it now absorbs the fiscal impact of the U.S. tax reform using distortionary taxes—Europe's welfare loss doubles to nearly 1.7 percent, as τ_c^* rises from 15.[illegible] to 18.4 percent. Thus, the benchmark case from Table 2 is an ideal scenario that gives Europe the privilege of responding to the budgetary effects of U.S. tax reforms using nondistortionary taxes. We showed that even in that case the global externalities are very large, and that relaxing this assumption only magnifies them.

The last important consideration regarding the tax-policy stance is to relax the assumption that foreign interest income is not taxed. This is critical for two reasons. First, the assumption is very unrealistic. The majority of industrial countries tax all individual capital income according to the residence principle [see Frenkel and Razin (1992) and Frenkel et al. (1992)], so income from foreign securities is taxed at similar rates as domestic capital income. Second, opening the economy to trade in financial assets implies that tax policy has a new margin of distortion—the wedge between the intertemporal marginal rate of substitution in consumption and the world interest rate.

The fourth row of Table 3 considers a case in which the United States applies the "full" residence principle to domestic capital and foreign interest incomes in the pre-tax-reform equilibrium (i.e., τ_k applies to both foreign interest and domestic capital incomes). Because we assumed discounted bonds, the easiest manner to introduce a foreign interest tax is via a tax on bond purchases τ_b, so that the amount of resources invested into foreign bonds in equation (2) becomes $(1 + \gamma)R_t(1 + \tau_b)B_{t+1}$. The value of τ_b consistent with the pre-reform value $\tau_k = 41.5$ percent is $\tau_b = 1.05$ percent. Note that, for reasons argued in Section I, subsection B, residence-based taxation and global trade in bonds require that τ_b be set at the same rate in both countries. In addition, although the tax on bonds distorts both the short- and long-run real interest rates, the reform we examine maintains τ_b unchanged. This ensures that there is a transitory increase in the world interest rate, while the long-run real interest rate is the same before and after the reform (as in all other experiments).

The results of the experiment with tax on bonds show that the distortion on the incentives for international borrowing alter the world distribution of the gains of the tax reform. The welfare effects now amount to a loss of 0.7 percent for the United States and a gain of 1.7 percent for Europe. Inspection of the transitional dynamics shows that the

[8] In Mendoza and Tesar (1995) we study the equilibria of tax-competition games under cooperative and noncooperative behavior that are triggered by the global externalities of tax reforms.

transitional increase in the rate of interest is larger when the tax on bonds is present (2.2 basis points versus 1.6 basis points in the benchmark case), and the increased cost of borrowing results in a smaller trade deficit and a larger decline in consumption for the United States along the transition path. The smaller trade deficit on impact reflects the fact that the U.S. economy is borrowing less, but the long-run trade surplus grows to 6.2 percent, compared to 3.8 percent without the tax on bonds. This occurs because the long-run cost of servicing the debt is higher with the tax on bonds—the long-run interest rate is now given by $r \approx \rho + \tau_b - \gamma\sigma$. In short, the tax on bonds weakens the smoothing effect and strengthens the income-redistribution effect, and since the former benefits the tax-reforming country and the latter the foreign country, there is a redistribution of welfare gains favoring Europe. Moreover, there is an additional fiscal effect because the larger transitional decline in consumption implies that the consumption tax must be increased more in order to satisfy the intertemporal budget constraint of the government—τ_c reaches 19.3 percent versus 14.8 percent without the tax on bonds.

If one reconsiders the labor tax reform in the presence of a tax on bonds, there is still a sizable redistribution of the welfare gains—U.S. welfare increases 2.8 percent instead of 3.1 percent, and European welfare increases 0.3 percent instead of falling 0.3 percent. In this case, however, the international transmission channels are weaker, and in fact contribute to make the U.S. tax reform sustainable by producing a welfare gain for European households without causing a welfare loss in the United States.

In addition to the overall tax-policy environment, the international implications of tax reforms depend on the structure of preferences and technology, in particular on the curvature of the cost-of-adjustment function, η, the leisure exponent, a, and the initial stock of foreign assets, b_0. The second panel of Table 3 conducts sensitivity analysis for these three key parameters in the case of the U.S. capital income tax reform.

First we examine the implications of sharply lowering η, to a value nearly 22 times smaller than in the benchmark. Predictably, [illegible] smoothing and income-redistribution ef[illegible] are much stronger, and the welfare gain (l[illegible] of the home (foreign) country is larger. W[illegible] adjustment costs are trivial, k and b are [illegible] perfect substitutes, and hence as the tax re[illegible] increases the post-tax domestic marginal pr[illegible] uct of capital, households borrow as much [illegible] necessary to enlarge k immediately so as [illegible] reset this marginal product at the level of [illegible] world interest rate. The latter rises more [illegible] before during the transition, because the h[illegible] country is relatively large in the world mark[illegible] but still the home country's trade deficit [illegible] share of GDP widens by more than 40 [illegible] centage points on impact, more than four ti[illegible] as in the benchmark. Note, however, that a[illegible] justment costs are needed only to [illegible] enough "sand in the wheels" of capital [illegible] cumulation on the margins represented by co[illegible] ditions (9)–(10), and that this is accompli[illegible] with adjustment costs that are trivial rel[illegible] to the size of the economy. In the benchm[illegible] case, U.S. adjustment costs converge rapi[illegible] and monotonically, to zero from a maxim[illegible] of only 0.8 percent of GDP on the date the [illegible] reform is introduced. We conclude, therefo[illegible] that while the positive effects of the tax refor[illegible] are highly dependent on the value of η, [illegible] U.S. (European) welfare gains (losses) r[illegible] between 2.9 and 3.8 (−0.8 and −1.8) for v[illegible] large variations in the curvature of adjustm[illegible] costs.

The case of $a = 0$, which effectively ma[illegible] labor supply inelastic at the level of the [illegible] endowment, is examined next (second row [illegible] the bottom panel of Table 3). This experim[illegible] demonstrates the importance of the intera[illegible] between consumption and leisure in the o[illegible] come of tax reforms. Setting $a = 0$ weak[illegible] both smoothing and income-redistribu[illegible] effects, as reflected in the smaller imp[illegible] and long-run changes in the U.S. tr[illegible] balance-GDP ratio relative to the bench[illegible] case. Despite these weaker effects, the U.S. [illegible] welfare gain is larger, and the European [illegible] smaller, than in the benchmark case. T[illegible] somewhat counterintuitive result follows f[illegible] two effects:

(1) Lowering a in the United States enl[illegible] the welfare gains of the tax reform for s[illegible] reasons as in a closed economy (see L[illegible]

1990). The system (12)–(15) implies that a lower a allows a given capital income tax cut to generate more output and consumption in the long run at a lower utility cost in terms of foregone leisure. By (12), the cut in τ_k increases k/y by an amount that is invariant to the value of a and is the same for closed and open economies. In the open economy, c/y falls both because the rise in k/y increases the investment-output ratio and because the tax cut leads to an increase in the net exports-GDP ratio to service the external debt. Given c/y, (15) implies that N is decreasing in a, so the long-run labor supply increases when a falls, and, given k/y, this results in larger output. Welfare increases because with the lower a the increase in consumption is valued more than the fall in leisure. In addition, the increased N and c provide a wider tax base, and hence imply that the increase in τ_c needed to maintain fiscal solvency is smaller.

(2) Lowering a^* along with a weakens the global spillovers of the tax reform because there can be no long-run outflow of physical capital away from Europe. Europe's long-run ratio k^*/y^* is pinned down by the foreign-country version of (12), so the ratio does not change because τ_k^* is unchanged, and with a Cobb-Douglas technology this ratio satisfies $k^*/y^* = (k^*/N^*)^{(1-\alpha)}$. Given that N^* is inelastic and k^*/y^* is constant, it follows that k^* and y^* are invariant to a U.S. tax reform. There is still a short-run capital outflow from Europe triggered by U.S. borrowing and temporarily higher world interest rates. Moreover, since with $a = 0$, output in both countries is predetermined on the date of the reform, the brunt of the adjustment required to generate the 7.9 European trade surplus on impact is absorbed by a larger transitional decline in European consumption.

The welfare impact of the weaker international transmission can be isolated by measuring the excess of the open-economy welfare gains over the closed-economy welfare gains for common values of a. With $a = 2.675$, the difference is nearly $^3/_4$ of a percentage point, while with $a = 0$ the difference narrows to 1/3 percentage points.

Interestingly, the dynamics of convergence to the balanced-growth path for European output and capital in the experiment with $a = 0$ are not monotonic, in sharp contrast to the benchmark case illustrated in Figure 2. Both European output and capital decline in the initial stages of the transition and then increase until they converge to their pre-tax-reform levels. Continuity implies that these nonmonotonic dynamics remain a feature of the model for low values of a. Thus, it is in principle possible to rationalize the observed nonmonotonicity of output and investment dynamics across countries (see Chari et al., 1996) as an outcome of the global externalities of tax-policy changes.

The last row of Table 3 examines the case in which the U.S. pre-tax-reform foreign asset position is zero, instead of the amount consistent with a 1-percent, long-run trade deficit as a fraction of GDP. This implies that the United States undertakes the tax reform starting from balanced trade, and is equivalent to reducing U.S. financial wealth from $b_0 = 0.62$ (which is consistent with $tb/y = -0.01$) to $b_0 = 0$. Since the present value of the household's expenditures must equal the present value of income plus financial wealth, the reduction in b_0 has an adverse effect on the U.S. welfare gain. The effect is quantitatively small because the initial bond positions and trade deficits in the benchmark case were small. However, since trade and foreign asset positions vary widely across countries and over time, and since the relevant value of b is the one prevailing *exactly* on the date of the tax reform (not an historical average), these results suggest that initial foreign asset positions are a key factor determining the outcome of tax reforms. For example, if the model is simulated assuming that $b_0 = 3$ instead of 0.62, which implies an initial U.S. trade deficit of 5 percent of GDP, the U.S. (European) net welfare gain (loss) increases (falls) to 3.5 (−1.3) percent.

Finally, it is worth noting that despite the sizable welfare gains produced by the tax reforms we examined, there still remains a very large social cost associated with the need to raise revenue through distortionary taxation. In a hypothetical scenario in which the current structure of distortionary taxes could be replaced with lump-sum taxation worldwide, the United States and Europe would make net welfare gains of nearly 13 and 20 percent, respectively.

III. Conclusions

This paper examines the positive and normative effects of tax reforms in the context of a two-country dynamic macroeconomic model with fully integrated capital and goods markets. International capital markets alter the response of the economy to tax reforms through two effects: (a) a *smoothing effect*, reflected in increased external borrowing to finance the expansion of the capital stock while smoothing consumption and leisure, and (b) an *international income-redistribution effect*, reflected in the long-run trade surplus needed to service the external debt accumulated for smoothing purposes. The model is calibrated to reflect the current stance of tax policy in Europe and the United States, assuming standard specifications for preferences and technology, and simulated to quantify the international implications of reforms replacing income taxes with higher consumption taxes. The model is solved numerically using a linear-approximation method augmented with shooting routines to determine solutions of competitive equilibria in which the world redistribution of financial wealth triggered by tax reforms is properly quantified.

The smoothing and income-redistribution effects imply that measures of the normative and positive implications of tax reforms are markedly different in open economies than in closed economies. The net welfare gains of eliminating U.S. income taxes can be up to 34 percent larger in the open-economy model and the cost of the macroeconomic transitional dynamics can be cut by more than half. Also, the excess welfare gain in the United States is obtained at the expense of a fall in European welfare. The enhanced efficiency of the U.S. economy, resulting from the reduction in tax distortions, leads European households to reallocate their savings from their own capital stock to international bonds and to sacrifice consumption and leisure during the transition. They recover only a fraction of the welfare cost of these adjustments via the income-redistribution effect.

Further analysis reveals that knowledge of the complete structure of direct and indirect taxes across countries is critical for assessing the outcome of tax reforms. In particular, taxes on interest income from external assets add a new margin of distortion by driving a wedge between the world real interest rate and the intertemporal marginal rate of substitution in consumption. Through this wedge, taxation of foreign interest income weakens the smoothing effects and strengthens the income-redistribution effect. As a result, a country undertaking a unilateral tax reform can transfer the benefits of the reform to a foreign country with unchanged taxes.

Our quantitative investigation demonstrates that unilateral tax reforms induce significant externalities on the rest of the world, which would lead fiscal authorities to engage in potentially damaging international tax competition. We do not study tax competition, but we do identify simple worldwide tax reforms yielding Pareto-efficient outcomes that leave households worldwide better off than under present tax structures. Moreover, the numerical methods developed here provide the basic ingredients for computing solutions of international tax-competition games, which we are producing in work in progress (see Mendoza and Tesar, 1995). Further research on how our quantitative results are altered by distributional issues across generations and income groups, the introduction of stochastic shocks or policy uncertainty, and endogenous government expenditure choices, all of which have been found important in theoretical work, are worth pursuing.

The international effects of tax reforms are particularly sensitive to three elements of the model's structure. First, capital-adjustment costs affect significantly the quantitative features of transitional dynamics of macroeconomic aggregates, although they are less relevant for welfare implications. In contrast, the elasticity of labor supply is key for both positive and normative implications of the model. If labor is inelastic, a domestic tax reform cannot lead to a permanent outflow of physical capital from abroad, and hence the smoothing and income-redistribution effects are sharply weakened. The third key parameter is the net foreign asset position of the reforming nation on the date the reform is implemented. A large net creditor in global markets has a stock of financial wealth that can be depleted to smooth consumption and leisure, without causing long-run income redistribution on account of permanently higher payments to service foreign debt.

Finally, some of our findings have interesting implications for other aspects of open-economy policy analysis that have been the focus of recent debate. First, tax reform produces at the outset large and persistent trade and fiscal deficits for the reforming nation, which are sustainable in the sense of being consistent with long-run solvency. Hence, large fiscal and external imbalances in countries embarked in far-reaching economic reforms can be equilibrium outcomes, and aiming to reduce them can be highly undesirable. Second, nonmonotonic convergence of macroeconomic aggregates to balanced-growth paths, which violate the predictions of the conventional closed-economy, balanced-growth framework, can emerge as a result of the international externalities induced by changes in economic distortions of particular nations.

REFERENCES

Aiyagari, S. Rao. "Optimal Capital Income Taxation with Incomplete Markets, Borrowing Constraints, and Constant Discounting." *Journal of Political Economy*, December 1995, *103*(6), pp. 1158–75.

Barro, Robert J. and Sala-i-Martin, Xavier. *Economic growth*. New York: McGraw-Hill, 1995.

Chari, V. V.; Christiano, Lawrence J. and Kehoe, Patrick J. "Optimal Fiscal Policy in a Business Cycle Model." *Journal of Political Economy*, August 1994, *102*(4), pp. 617–52.

Chari, V. V.; Kehoe, Patrick J. and McGrattan, Ellen. "The Poverty of Nations." Discussion paper, Research Department, Federal Reserve Bank of Minneapolis, 1996.

Cooley, Thomas F., ed. *Frontiers of business cycle research*. Princeton, NJ: Princeton University Press, 1995.

Cooley, Thomas F. and Hansen, Gary D. "Tax Distortions in a Neoclassical Monetary Economy." *Journal of Economic Theory*, December 1992, *58*(2), pp. 290–316.

Frenkel, Jacob A. and Razin, Assaf. *Fiscal policies and the world economy: An intertemporal approach*, 2nd Ed. Cambridge, MA: MIT Press, 1992.

Frenkel, Jacob A.; Razin, Assaf and Sadka, Efraim. *International taxation in an integrated world*. Cambridge, MA: MIT Press, 1991.

Greenwood, Jeremy and Huffman, Gregory W. "Tax Analysis in a Real-Business-Cycle Model." *Journal of Monetary Economics*, April 1991, *27*(2), pp. 167–90.

Greenwood, Jeremy and Kimbrough, Kent P. "Capital Controls and Fiscal Policy in the World Economy." *Canadian Journal of Economics*, November 1985, *18*(4), pp. 743–65.

King, Robert G. and Rebelo, Sergio T. "Public Policy and Economic Growth: Developing Neoclassical Implications." *Journal of Political Economy*, October 1990, Part 2, *98*(5), pp. S126–50.

Lipton, David; Poterba, James; Sachs, Jeffrey and Summers, Lawrence. "Multiple Shooting in Rational Expectations Models." *Econometrica*, September 1982, *50*(5), pp. 1329–33.

Lipton, David and Sachs, Jeffrey. "Accumulation and Growth in a Two-Country Model: A Simulation Approach." *Journal of International Economics*, August 1983, *15*(1/2), pp. 135–59.

Lucas, Robert E., Jr. *Models of business cycles*. Oxford: Blackwell, 1987.

________. "Supply-Side Economics: An Analytical Review." *Oxford Economic Papers*, April 1990, *42*(2), pp. 293–316.

Mendoza, Enrique G. "Real Business Cycles in a Small Open Economy." *American Economic Review*, September 1991, *81*(4), pp. 797–818.

Mendoza, Enrique G. and Tesar, Linda L. "Supply-Side Economics in a Global Economy." National Bureau of Economic Research (Cambridge MA) Working Paper No. 5096, April 1995.

Mendoza, Enrique G.; Tesar, Linda L. and Razin, Assaf. "Effective Tax Rates in Macroeconomics: Cross Country Estimates of Tax Rates on Factor Incomes and Consumption." *Journal of Monetary Economics*, December 1994, *34*(3), pp. 297–323.

Persson, Torsten and Tabellini, Guido. "Double-Edged Incentives: Institutions and Policy Coordination," in Gene Grossman and Kenneth Rogoff, eds., *Handbook of international economics*, Vol. III, New York: North-Holland, 1995.

Pigou, A. C. *A study in public finance*. London: Macmillan, 1947.

Currencies and the Allocation of Risk: The Welfare Effects of a Monetary Union

By PABLO ANDRÉS NEUMEYER*

In a general equilibrium model with incomplete asset markets, nominal securities, and mean-variance preferences, a monetary union is desirable when the gain from eliminating excess volatility of nominal variables exceeds the cost of reducing the number of currencies with which to hedge risks. (*JEL* F33, D52, E44, G10)

At an informal level it is often argued, especially among businessmen and politicians, that reducing uncertainty is an important benefit of a monetary union. At the heart of this argument lies the implicit assumption that some risks are uninsurable. Most of the literature on the welfare effects of exchange rate regimes and currency unions, however, does not explicitly address the connection between the exchange rate regime and the incompleteness of financial markets.[1] This article applies general equilibrium theory to investigate how a monetary union affects the efficiency of the allocation of risks in an economy with incomplete asset markets.

The paper compares a monetary union with exchange rate regimes that fall short of permanently fixing the exchange rate (including floating exchange rates, fixed exchange rates subject to realignments, and exchange rate bands). It is shown that the adoption of a monetary union involves a trade-off between the benefits of reducing "excessive" exchange rate risk and the costs of reducing the number of assets in the economy.

In spite of the general belief that "excessive" exchange rate variability harms the economy, proving this in a formal setting has been difficult. The difficulty arises because fluctuations in exchange rates that reflect economic shocks (income and preference shocks in this paper) actually *help* to allocate resources efficiently. The paper shows that the exchange rate risks that are *not* caused by economic shocks reduce the efficiency of financial markets. A potentially important source of this type of excess exchange rate risk appears when political factors influence monetary policy. Political interference in monetary affairs implies that given the realization of an economic shock, there still is uncertainty about the future actions of monetary authorities if such actions will be influenced by future political events. For example, the timing of monetary stabilizations in inflationary economies, the value at which currencies enter a fixed exchange rate regime, and exchange rate

* Department of Economics, University of Southern California, Los Angeles, CA 90089. I want to thank the referees for helpful suggestions. Comments from Caroline Betts, Alberto Bisin, Matthew Canzoneri, Pierre André Chiappori, Tim Kehoe, Thorsten Hens, Michael Magill, Juan Pablo Nicolini, Maurice Obstfled, Gerardo della Paolera, Heraklis Polemarchakis, Martine Quinzii, Peter Rosendorff, Federico Sturzenegger, Pedro Teles, Carlos Vegh, Makoto Yano, and seminar participants at Boston College, the University of California-Berkeley, the University of California-Davis, UCLA, the University of California-Santa Cruz, Georgetown University, Stanford University, the International Monetary Fund, DELTA (Paris), the Universidad Carlos III de Madrid, the Universidad T. Di Tella (Buenos Aires), the Seminaire d'Economie Monetaire International in Paris, the SEDC meetings in Barcelona, and the Econometric Society 7th World Congress are gratefully acknowledged. Part of the research contained in this paper was done while I was visiting the Universidad T. di Tella, DELTA, and the Universidad Carlos III de Madrid. I am grateful to these institutions for their hospitality and to the Commission of European Communities for financial support (Grants ERBCHBICT930544 and ERBCHBGCT 920146). Any errors are my own.

[1] See Robert A. Mundell (1973) and Elhanan Helpman and Assaf Razin (1982) for early analyses of the relation between the exchange rate regime and the efficient sharing of international risks.

ments are instances of monetary policy decisions that depend on the realization of political shocks.[2] In such cases the expected variability of nominal variables is too high due to their dependence on future political shocks. These excess fluctuations in price levels are socially costly because they contaminate the real payoffs of nominal financial contracts, reducing their ability to help agents hedge against economic shocks. Currency unions and permanently fixed exchange rate regimes can be viewed as monetary rules that attempt to improve welfare by insulating money from domestic politics.[3,4]

The adoption of a fixed exchange rate regime or a currency union also has costs. When fluctuations in the value of money reflect economic shocks, some exchange rate variability is "good" because, by making the real payoffs of nominal assets denominated in different currencies distinct, it increases the insurance opportunities available through trade in nominal assets. The loss of monetary independence entailed by fixed exchange rate regimes, or currency unions, is socially costly because it makes the real payoff of assets denominated in different currencies equivalent, effectively reducing the number of financial instruments with which economic agents can share risks.

The main result of the paper is that switching from a monetary regime with national central banks to a currency union increases welfare when the gain from eliminating excess monetary volatility exceeds the cost of reducing the number of financial instruments in the economy. This case for a monetary union has not been previously modeled in the international finance literature.

The modern theory of general equilibrium with incomplete financial markets provides a natural and fruitful framework in which to study the effect of monetary and exchange rate regimes on the allocation of risks (see, for example, Michael Magill and Martine Quinzii, 1996). Helpman (1981), John H. Kareken and Neil Wallace (1981), and Robert E. Lucas, Jr. (1982) were the first to use general equilibrium models to evaluate the welfare effects of exchange rate regimes. These seminal contributions prove welfare equivalence theorems for alternative exchange rate regimes in environments where asset markets are complete and money is neutral.[5] Following these contributions, the connection between the exchange rate regime and the efficiency with which an incomplete set of nominal financial markets can allocate risk was studied by Helpman and Razin (1982). In their model, a floating exchange rate regime dominates a fixed one because the latter effectively reduces the number of assets in the economy. Helpman and Razin felt that there ought to be conditions under which a fixed exchange rate regime outperforms a floating one, but as they could not find any, they left the question open for future research. This paper takes up the task of comparing alternative exchange rate regimes where Helpman and Razin left it, with the aid of two ideas borrowed from the modern theory of general equilibrium with incomplete financial markets.

The first insight of this literature for the purpose of this paper is the geometric approach to the role of money in economies with incomplete markets (Magill and Quinzii, 1992).

[3] On the timing of stabilizations, Domingo F. Cavallo and Joaquín A. Cottani (1997 p. 17) write "After 18 months of political hesitation, the Menem government finally decided to tackle the problem at its root and launched the Convertibility Plan." See also Alberto Alesina and Allan Drazen (1991) for the timing of stabilizations. Milton Friedman and Anna Schwartz (1963) discuss the politics underlying the resumption of gold and the debate on bimetallism after the American Civil War. Jeffry A. Frieden (1994a) studies the politics of currency crises in the European Monetary System (ERM) and Barry Eichengreen (1992) analyzes the politics of the gold standard.

[4] For example, current plans to form a European currency union include the creation of a supranational central bank designed to prevent monetary policy from being subject to the influence of domestic political authorities (Article 107, Maastricht Treaty; see Charles Bean, 1992). For a discussion of the political and institutional aspects of a monetary union, see Michael Mussa (1997).

[5] The literature following these contributions can be classified according to the frictions generating the monetary nonneutralities that make the exchange rate regime relevant. In this paper nonneutrality arises because asset markets are incomplete and some financial contracts are written in nominal terms. Other frictions are price rigidities (Mundell, 1961) and the use of seignorage as a tax (Matthew B. Canzoneri and Carol Ann Rogers, 1990).

Money matters because it changes the reallocations of income across states of nature that are attainable through trade in assets (it tilts the vector subspace of income transfers spanned by the matrix of asset payoffs).[6] This idea serves to clarify the result in Helpman and Razin and is a central element of this study. The second idea borrowed from the theory of incomplete markets is that future price levels can depend, not only on shocks to preferences or endowments (economic shocks), but also on other sources of uncertainty called "sunspots" in the general equilibrium literature [see Cass (1989) and Paolo Siconolfi (1991)]. These sunspots are the political shocks that influence monetary policy in this paper and create the opening (that is missing in Helpman and Razin) for making a new case for fixed exchange rate regimes.

The remainder of the paper is organized as follows: Section I describes the model; Section II characterizes the equilibrium geometrically; Section III establishes the welfare costs of monetary politics; Section IV evaluates the welfare effects of a monetary union; and Section V concludes.

I. Description of the Model

Consider a two-date general equilibrium model of an exchange economy with one perishable commodity. The world's population is represented by the set $\mathbf{H} = \{1, \dots, H\}$[7] and individuals are indexed by h. The partition of the set $\mathbf{H}$ into countries defines I subsets of individuals, $\mathbf{H}_i$. These countries are indexed by i with $i \in \mathbf{I} = \{1, \dots, I\}$.

Economic activity extends over two dates. Agents know the present, called state 0, and face an uncertain future. At the second date 1, one of S possible states is realized. Each of the states at date 1 is defined by the realization of two types of shocks: *economic* and *political.* The *economic* uncertainty is associated with the realization of each individual's income. The states of nature that depend on the realization of this type of uncertainty will be indexed by σ and take values in the set $\Sigma = \{1, \dots, \Sigma\}$; $\Sigma > 1$. The *political* factors that influence monetary policy will be represented by the random variable, θ, taking values in the set $\mathbf{\Theta} = \{1, \dots, \Theta\}$. It is possible to write the set $\mathbf{\Theta}$ as $\mathbf{\Theta} = \mathbf{\Theta}_1 \times \mathbf{\Theta}_2 \times \dots \times \mathbf{\Theta}_I$ and its elements as $\theta = (\theta_1, \theta_2, \dots, \theta_I)$, interpreting each element of the vector θ as the political uncertainty that is generated in country i. Summarizing, a state at date 1 is $s = (\sigma, \theta)$ and the state space of the economy is $\{0\} \cup \mathbf{S}$, where $\mathbf{S} = \mathbf{\Sigma} \times \mathbf{\Theta}$. The probability of state s occurring at date 1 is denoted by $\pi(s)$.

In each state the quantity of money that is injected in each country is ${}_iM(s)$. Without loss of generality, it is assumed that for all countries ${}_iM(0) = 1$. Monetary policy in country i is represented by a function ${}_iM: \mathbf{S} \to \mathbb{R}^S_{++}$. An independent central bank with no discretionary power will follow a rule which is independent of the political shocks.[8,9]

[6] The idea that changes in the price level have real effects because they alter the space of income transfers across states of nature that are attainable through trade in assets was originally developed in David Cass (1985), Yves Balasko and Cass (1989), and John Geanakoplos and Andreu Mas-Colell (1989). See Heracles Polemarchakis (1988) for economies with multiple currencies. At the time when Helpman and Razin wrote their paper, the geometry of monetary policy in economies with incomplete markets was not well understood. This idea helps to understand not only Helpman and Razin but, also, other important articles in international macroeconomics such as Torsten Persson and Lars E. O. Svensson (1989) and Svensson (1989).

[7] *Notation:* Bold upper-case letters denote sets, upper-case italic letters denote the number of elements in a set, and lower-case italic letters denote an element of the set. For example, $\mathbf{H}$ is the set of individuals, H is the number of individuals, and h is one particular individual.

[8] The money supply function ${}_iM$ is consistent with empirical literature on monetary policy shocks [see Lawrence J. Christiano and Martin Eichenbaum (19..) for a survey] and with many computational experiments where researchers work with monetary policy rules of the form $M = f(\sigma) + \varepsilon$, where $f(\sigma)$ is a feedback rule that depends on the state of the economy and ε is a random shock, that can be interpreted as a function of the political shocks.

[9] The political shocks also can be interpreted as self-fulfilling beliefs about the future actions of a benevolent monetary authority [Guillermo A. Calvo (1988); V. V. Chari et al. (1996)]. Under a discretionary regime, policy makers can be pushed into pursuing an inflationary policy when the private sector, for whatever reason, expects inflation, and monetary authorities find it too costly

It is possible to interpret the political shocks as uncertainty about the preferences of a benevolent policy maker that maximizes a weighted average of the utility attained by the residents under its jurisdiction at a competitive equilibrium. This uncertainty can arise because private agents do not know the policy maker's preferences or because they do not know who will be in power in the future.

Every individual has an endowment vector $\omega^h \in \Omega$, where

$$\Omega = \{\omega^h \in \mathbb{R}^{S+1}_{++} \mid \forall \sigma \in \Sigma : \omega^h(\sigma, \theta) = \omega^h(\sigma, \theta') \text{ for every } \theta, \theta' \in \Theta\}.$$

The definition of the set Ω formalizes the notion that political shocks are distinct from economic ones. Notice, however, that the two types of shocks do not need to be statistically independent.

Preferences are not stochastic and are represented by the linear-quadratic utility function

$$(1) \quad u^h = x^h(0) + \sum_{s \in S} \pi(s)[cx^h(s) - \tfrac{1}{2}x^h(s)^2],$$

that is defined over the consumption set $\mathbb{R}^{S+1}$, and where $x^h(s)$ is consumption in state s. In order to guarantee monotonicity, it is assumed that $c > \Sigma_{h \in H}\, \omega^h(s)$ for all $s \in \mathbf{S}$.[10] This specification of preferences specializes the model to the capital asset pricing model (CAPM). It has the advantage of yielding a closed-form solution and of conferring very nice geometrical properties to the equilibrium.

The securities that are traded in world financial markets are an indexed bond that pays a unit of the consumption good in every state, and a nominal bond denominated in each of the currencies.[11] The nominal bonds pay one unit of money in every state; their real payoffs are the inverse of the price level in currency i. The payoffs of all the securities can be collected in an $S \times J$ matrix of real asset payoffs,

$$A(p) = \begin{bmatrix} 1 & \dfrac{1}{{}_1p(1)} & \cdots & \dfrac{1}{{}_Ip(1)} \\ \vdots & \vdots & & \vdots \\ 1 & \dfrac{1}{{}_1p(s)} & \cdots & \dfrac{1}{{}_Ip(s)} \end{bmatrix},$$

where $p = [{}_1p, \ldots, {}_Ip]$. The column vectors of A, a_j, represent the payoff of asset j. The row vectors of A, $a(s)$, represent the payoff of all the securities in state s.

As there are no trading frictions associated with trade in either assets or commodities, the law of one price must hold, i.e.,

$$(2) \quad e_{ij}(0)_i q_k = {}_j q_k \qquad \text{for every } i, j, \text{ and } k,$$

$$e_{ij}(s)_i p(s) = {}_j p(s) \qquad \text{for every } i, j, \text{ and } s \in \{0\} \cup \mathbf{S},$$

where e_{ji} is the price of currency j in terms of currency i, ${}_jp$ is the price level in currency j, and ${}_jq_k$ is the price of security k in currency j.

These no-arbitrage conditions imply that the date 0 budget constraint can be written as:

$$x^h(0) + qy^h = \omega^h(0),$$

where $q \in \mathbb{R}^J$ is the vector of real asset prices and $y^h \in \mathbb{R}^J$ denotes the portfolio of individual h, with y^h_j representing holdings of security j.

... in terms of output, or in terms of the need to use distortionary taxes) not to accommodate these ...

[10] Allowing c to differ across individuals does not alter the paper's results but adds to its notational complexity.

[11] This asset structure implicitly includes futures currency contracts. The absence of arbitrage opportunities in financial markets implies that a futures contract to exchange currency i for currency j is equivalent to being long in the nominal bond denominated in currency j and short in the bond in currency i.

The date 0 budget constraint states that consumption plus the value of the securities purchased at date 0 must equal each individuals endowment at date 0.

At the second date, agents discover their endowment $\omega^h(s)$, and deliveries on the financial promises made at date 0 are carried out. The budget constraints are

$$x^h(s) = \omega^h(s) + a(s)y^h$$

in each of the S states at date 1.

Summarizing, the consumer's budget set is

$$(3)\quad \mathcal{B}^h(p, q) = \left\{ (x^h, y^h) \in \mathbb{R}^S \times \mathbb{R}^J : x^h - \omega^h = \begin{pmatrix} -q \\ A(p) \end{pmatrix} y^h \right\}.$$

The money markets clear if, for $s = 0, 1, \ldots, S$,

$$(4)\qquad {}_iM(s) = {}_ip(s)_i\omega(s),$$

where ${}_i\omega(s)$ is the aggregate endowment of country i in state s. These money market-clearing conditions are quantity equations with unitary velocity.[12]

Financial markets clear if

$$\sum_{h \in \mathbf{H}} y^h = 0.$$

The economy is in *equilibrium* when individuals choose their consumption and portfolios in order to maximize their utility in the budget set $\mathcal{B}^h(p, q)$, the money market-clearing conditions (4) hold, and asset markets clear. Walras Law implies that the good ket clears as well.

II. The Geometry of Equilibrium

An immediate consequence of the market-clearing conditions is that the rium matrix of asset payoffs is

$$A(p^*) = \begin{pmatrix} 1 & \dfrac{{}_1\omega(1)}{{}_1M(1)} & \cdots & \dfrac{{}_I\omega(1)}{{}_IM(1)} \\ \vdots & \vdots & & \vdots \\ 1 & \dfrac{{}_1\omega(S)}{{}_1M(S)} & & \dfrac{{}_I\omega(S)}{{}_IM(S)} \end{pmatrix}$$

where p^* denotes the equilibrium price. The convention that a starred variable cates its equilibrium value will be adopt the remainder of the paper.

Observing this matrix sheds some lig the problem of understanding the reasons hind the existence of securities written in inal terms. When they are evaluated in terms, nominal securities are a very sop cated set of financial contracts that are con gent on whatever determines the value money. If securities are standardized countries some exchange rate variability makes price levels in different countries tinct) is "good" because it increases the surance opportunities available through in nominal assets. Observe that for any monetary policy the matrix $A(p^*)$ will full rank for almost any configuration of dowments. On the other hand, the loss of etary independence entailed by fixed exch rates is costly: fixed exchange rates make money supply endogenous and the price in different countries colinear, inducing of rank in $A(p)$.[13]

The budget constraints (3) imply th excess demand of each individual at date 1 to belong to the linear subspace spanned

[12] The monetary institutions underlying this model are the extension to an international setting of those presented in Magill and Quinzii (1992). Money does not appear in the budget constraints because it is assumed that at the beginning of each period consumers sell their endowments to the central bank, then they trade in asset and currency markets, and at the end of each period they sell their money holdings back to the central banks in exchange for consumption goods. See the working paper version (1995) of this article for details.

[13] The drop of rank of the matrix of asset pay duced by a fixed exchange rate regime is the reason a floating exchange rate regime is better than a fix in Helpman and Razin (1982).

the columns of $A(p^*)$. This hyperplane is called the marketed subspace (of $\mathbb{R}^S$) and its dimension is equal to the number of nonredundant assets in the economy (the independent columns of A), J. When asset markets are incomplete (i.e., $J < S$) the marketed subspace has a lower dimension than the consumption set at date 1, and, in general, the excess demand that corresponds to an equilibrium in an economy with complete markets will not be feasible. Equilibrium excess demands in an economy with incomplete markets will be the closest point in the marketed subspace to the complete markets allocation. Monetary policy matters when changes in M alter the marketed subspace.

The complete markets allocation is given by[14]

$$\hat{x}(\omega) = \begin{pmatrix} \omega^h(0) + \dfrac{1}{1+r^*} E\left(\omega_1^h - \dfrac{\omega_1}{H}\right) \\ - \operatorname{cov}\left[\left(\dfrac{\omega_1}{H}\right), \left(\omega_1^h - \dfrac{\omega_1}{H}\right)\right] \\ \dfrac{\omega_1}{H} \end{pmatrix},$$

where $1/(1 + r^*)$ is the price at date 0 of a riskless unit of the consumption good at date 1. This expression shows that when asset markets are complete there is perfect risk sharing, or full insurance. This can be readily derived from the fact that under this allocation of resources, each individual consumes the average world output at date 1, i.e., $\hat{x}_1^h = \omega_1/H$ for all h.

The assumption about the functional form of the utility function implies that the utility that individuals attain at an equilibrium can be expressed as

$$(5)\quad u^h(x^{*h}) = u^h(\omega^h) + d(x_1^{*h}, \omega_1^h)$$
$$= u^h(\hat{x}^h) - d(\hat{x}^h, x_1^{*h}),$$

where $d(x, y) = \Sigma\, \pi(s)(x(s) - y(s))^2$ measures the distance between the points x and y in $\mathbb{R}^S$. The distance between the date 1 endowment and equilibrium consumption is an exact measure of the gains from trade in assets. The distance between the date 1 complete markets allocation and equilibrium consumption measures the inefficiency introduced by the incompleteness of financial markets. Equilibrium portfolios pick the excess demand that minimizes the distance between the complete markets excess demand and the marketed subspace.[15]

Figure 1 illustrates the geometry of equilibrium allocations in an economy with two assets and three states of nature at date 1. The graph is drawn in the space of date 1 consumption: a three dimensional space where each axis represents consumption in one of the states. If there is no trade in assets the agent can only consume at the endowment point, ω^h. Financial markets allow individuals to reallocate income across states of nature. However, the budget set, (3), constrains individual excess demands to lie in the marketed subspace. In this example, the marketed subspace is the plane in $\mathbb{R}^3$ that contains the endowment point and is spanned by the payoff vectors of assets a_1 and a_2. Figure 1 shows that equilibrium consumption at date 1, x_1^{*h}, is the closest point in the marketed subspace to the complete markets allocation, $\hat{x}_1^h$. The individual depicted in the figure attains this allocation with a long position in asset a_1 and a short position in asset a_2.

Monetary policy has real effects because price levels determine the position of the payoff vectors of assets a_1 and a_2 and, hence, that of the marketed subspace. For example, if the price level corresponding to the currency in which asset a_1 is denominated increases in state 3, then the payoff of asset a_1 in state 3 (vertical axis) will be smaller, and the marketed subspace will tilt down. The equilibrium consumption of agent h will change because it will cease to be feasible. With the new price-

[14] The complete markets allocation can be easily computed by replacing the payoff matrix A by the identity in expression (A2) in the Appendix.

[15] Technically, the equilibrium excess demand with incomplete markets is the π-orthogonal projection of the complete markets excess demand onto the marketed subspace.

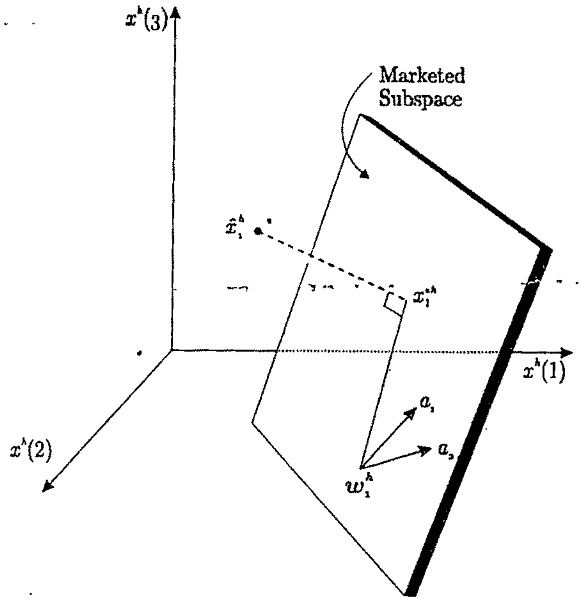


FIGURE 1. THE GEOMETRY OF EQUILIBRIUM

level vector, equilibrium consumption will be the closest point to the complete markets allocation lying in the new marketed subspace. Due to the way in which the plane was tilted, it will be closer to the autarky point and it will be further away from the complete markets allocation. Individual h, a lender in currency 1, will be worse off.

If country 2 pegs its currency to country 1, the payoff vectors corresponding to the two assets will become colinear, and individuals will only be able to consume on a point on the line that crosses their endowment in the direction of asset 1. In general, this will lower the utility of h.

III. The Welfare Cost of Monetary Politics

The objective of this section is to formalize the idea that the *excessive* nominal (exchange rate and price level) volatility that arises from political shocks reduces the gains from trading assets and, hence, welfare. This creates the opening for making a new case for fixed exchange rates in countries where this volatility is large.

In order to analyze the welfare costs of contaminating the real payoff of nominal assets with political shocks, equilibrium allocations are compared with the ones that would obtain in an hypothetical economy where monetary policy is made independent of θ, by replacing it by its expected value conditional on σ. Let $_i\bar{M} \in \mathbb{R}^S$ be a vector with elements $_i\bar{M}(\sigma) = \{\Sigma_\theta\ \pi(\theta|\sigma)[1/_iM(\sigma,\ \theta)]\}^{-1}$ and $\bar{M} = [_1\bar{M}, \ldots, {_I\bar{M}}] \in \mathbb{R}^S$ represent monetary policy in the I countries.

The utility gains that each individual can derive from eliminating excess nominal variability can be exactly measured. Recall that equilibrium utilities can be written as $u^h(\omega^h) + d(x_1^{*h},\ \omega_1^h)$. Let $\bar{u}^h$ and $\bar{x}_1^{*h}$ be equilibrium utility and consumption under monetary policy $\bar{M}$, respectively. It follows that the difference $\bar{u}^h - u^h$ is equal to $d(\bar{x}_1^{*h}, \omega_1^h) - d(x_1^{*h}, \omega_1^h)$, i.e., the difference in the gains from trade in assets in the two economies. The computation of this expression (shown in the Appendix) reveals that the dependence of monetary policy on political shocks reduces the gains from trade in assets:

$$\text{(6)}\quad \bar{u}^h - u^h = z^{h\prime}[(\bar{A}'\Lambda\bar{A})^{-1} - (A'\Lambda A)^{-1}]z^h \geq 0$$

where $z^h = A\Lambda(\pi)[(\omega_1/H) - \omega_1^h]$ and the matrix $\bar{A}$ corresponds to monetary policy $\bar{M}$. The matrices $A'\Lambda A$ and $\bar{A}'\Lambda\bar{A}$ are the variance-covariance matrices of asset returns with monetary policies M and $\bar{M}$, respectively. The difference is the *excess volatility* introduced by the political uncertainty, which is a semidefinite matrix.

In a world economy where in at least one country monetary policy depends on political shocks, individuals are diverse ($z^h \neq 0$ for some h), and asset markets are incomplete, the inequality in (6) will be strict for at least one agent.

PROPOSITION 1: *Let there be more than two agents in each country and fewer assets than states of the world. Then it is generically true that any monetary policy that is sensitive to political shocks supports an equilibrium that is Pareto inferior to the one corresponding*

... to some other monetary policy without ... shocks.

(See the Appendix for a formal statement of the proposition and its proof.)

For any monetary policy that introduces political uncertainty in the economy, there exists another policy, that is independent of θ, and is Pareto superior.

The intuition underlying Proposition 1 and expression (6) is best seen in the simple case where the variance-covariance matrices are diagonal, i.e., $\mathrm{cov}[(1/_i p^*), (1/_j p^*)] = 0$ for all i, j. In this case, the expression for the welfare gain of eliminating the monetary politics reduces to[16]

$$\text{(7)}\quad \bar{u}^h - u^h = \sum_i \frac{\left[\mathrm{cov}\left(e^h, \dfrac{\omega_i}{_i\bar{M}}\right)\right]^2}{\mathrm{var}\left(\dfrac{\omega_i}{_i\bar{M}}\right)} \times \frac{\sum_\sigma \pi(\sigma)\,\mathrm{var}\left(\dfrac{\omega_i}{_iM}\bigg|\sigma\right)}{\mathrm{var}\left(\dfrac{\omega_i}{_iM}\right)},$$

where $e^h \equiv \omega_1/H - \omega_1^h$ and $\mathrm{var}[(\omega_i/_iM)|\sigma] = \Sigma_\theta \pi(\theta|\sigma)[\omega_i(\sigma)/_iM(\sigma,\theta) - \omega_i(\sigma)/_i\bar{M}(\sigma)]^2$.

Expression (7) illustrates that any monetary policy for which $M(\sigma, \theta) \neq M(\sigma, \theta')$ is dominated by policy $\bar{M}$. The term $\Sigma_\sigma \pi(\sigma)\mathrm{var}[(\omega_i/_iM)|\sigma]$ is the expected value of the variance of national price levels conditioned on σ; it measures the sensitivity of monetary policy to political uncertainty. It is easy to see in (7) that the gains from eliminating excess nominal variability is an increasing function of this variable. If one thinks that there is a country zero, where the price level is constant and equal to one, $\omega_i/_iM$ would be the exchange rate between currency i and currency 0 and the term $\Sigma_\sigma \pi(\sigma)\mathrm{var}[(\omega_i/_iM)|\sigma]$ would be the excess exchange rate volatility.[17]

Equation (7) can also be written as

$$\bar{u}^h - u^h = \sum_i (y_i^{*h})^2 \left(1 + \frac{\Sigma_\sigma \pi(\sigma)\,\mathrm{var}\left(\dfrac{\omega_i}{_iM}\bigg|\sigma\right)}{\mathrm{var}\left(\dfrac{\omega_i}{_i\bar{M}}\right)}\right) \times \sum_\sigma \pi(\sigma)\,\mathrm{var}\left(\frac{\omega_i}{_iM}\bigg|\sigma\right).$$

The welfare loss that traders suffer because of the excess nominal variability in currency i is proportional to the square of their equilibrium trading position in that currency.[18]

Proposition 1 contributes to the general equilibrium literature on the welfare costs of monetary instability by showing that excessive monetary *variability* is socially costly because tampering with nominal financial contracts has negative effects on *financial* markets. This welfare cost of monetary instability forms the basis for Irving Fisher's (1936) case in favor of monetary rules. Friedman's (1960, 1962) case for rules rested, among other things, on his view that "(...) it is essential to prevent monetary policy from being a day-to-day plaything at the mercy of every whim of current political

[16] If $\mathrm{var}(\omega_i/_i\bar{M}) = 0$ and $\Sigma_s \pi(s)\mathrm{var}[(\omega_i/_iM)|s] > 0$, the bond denominated in currency i will not be traded and the right-hand side of equation (7) will be zero. If $\mathrm{var}(\omega_i/_iM) = 0$ the bond in currency i will be a perfect substitute for the real bond and the right-hand side of equation (7) will also be zero.

[17] If the value of the reference currency has a positive variance, then the relation between the variability of exchange rates and price levels is given by the formula

$$\frac{\mathrm{var}(e_i)}{E(e_i)^2} \simeq \frac{\mathrm{var}(p_i)}{E(p_i)^2} + \frac{\mathrm{var}(p_r)}{E(p_r)^2} - 2\frac{\mathrm{cov}(p_i, p_r)}{E(p_i)E(p_r)},$$

where p_r is the price level of the reference currency.

[18] This expression provides a theoretical basis for Frieden's (1994b) empirical observation that the special interest groups that exert the strongest political pressure to reduce exchange rate variability are those linked to the financial sector.

authorities.'' (Friedman, 1962 p. 224). Thus, Proposition 1 can be interpreted in terms of the literature that studies whether a monetary authority should follow preannounced *rules* or whether it should be granted *discretionary* power to conduct policy. A fixed exchange rate regime can be viewed as a monetary rule that is socially desirable when it removes excessive exchange rate variability. This case for a monetary rule is different from Finn E. Kydland and Edward C. Prescott's (1977) because their case for a rule is based on the gains from reducing the level of inflation, not its variability.

IV. The Welfare Effects of a Monetary Union

A monetary union entails the elimination of the U currencies of the countries that form the union and the creation of a common currency with its associated monetary policy. The disappearance of nominal assets in the merging currencies, and the creation of the new currency, transform the financial structure of the economy. The new asset structure is represented by the matrix

$$A_u(p) = \begin{bmatrix} 1 & \frac{1}{{}_1p(1)} & \cdots & \frac{1}{{}_{I-U}p(1)} & \frac{1}{{}_up(1)} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & \frac{1}{{}_1p(S)} & \cdots & \frac{1}{{}_{I-U}p(S)} & \frac{1}{{}_up(S)} \end{bmatrix},$$

where the equilibrium price level in the union will be determined by the quantity equations ${}_uM(s) = {}_up(s)\omega_u(s)$, and ${}_uM$, ${}_up$, and ω_u are the union's money supply, price level, and its aggregate endowment, respectively.

A regime with *permanently* fixed exchange rates in which countries peg their currencies to that of a ''center'' country (like the United States in the Bretton Woods system) is equivalent to a monetary union, with ${}_up$, ${}_uM$, and ${}_u\omega$ being the center country's price level, money supply, and aggregate income, respectively.

The objective of this section is to evaluate the welfare effects of a monetary union. To that extent, the effects of reducing exchange rate variability will be isolated from other effects on the asset structure.

The institutional design of the new central bank will determine the sensitivity of the union's monetary policy to political factors. The success of the new institutions, relative to the old regimes, in insulating monetary policy from political pressures can be measured by comparing the welfare costs of the excessive nominal instability under the two monetary regimes. Thus, the gains from switching to the union's monetary policy decision-making process will be $(\bar{u}_0^h - u_0^h) - (\bar{u}_u^h - u_u^h)$, where the subscripts 0 and u refer to the economies before and after the union, respectively. If the design of the union's central bank achieves the goal of divorcing money from politics, the difference $\bar{u}_u^h - u_u^h$ will be nil, and the gains from switching to the currency union will equal the costs of the excess nominal variability in the old regime. The case for a monetary union is strong when the latter is large.

The welfare effects of changing the number of assets in the economy can be studied by comparing the equilibrium utilities that can be attained in economies with the asset structures represented by $\bar{A}_0$ and $\bar{A}_u$, i.e., $\bar{u}_0$ and $\bar{u}_u$. Looking at the allocation generated by these matrices, instead of A_0 and A_u, is useful in separating the effect of changing the number of assets from the effect of changing the excess nominal volatility. The loss of nominal assets in the U disappearing currencies and the creation of assets in the union's currency reduces the dimension of the marketed subspace.

If the payoffs of the nominal bonds in the new currency are a linear combination of the payoffs of the nominal bonds in the disappearing currencies, then the marketed subspace generated by $\bar{A}_u$ will be included in the one generated by $\bar{A}_0$ and, generally, $\bar{u}_0 - \bar{u}_u$ will be positive. This will be the case if the money supply of the union and its members is constant, and in the case of a move toward a fixed exchange rate regime. In the more general case, when the new nominal bonds are not in the previous marketed subspace, the new marketed subspace will have a lower dimension but will also change its position in R^S, and $\bar{u}_0 - \bar{u}_u$ can be either positive or negative.

In terms of Figure 1, in the first case, the change in the asset structure of the economy

replaces the plane, spanned by the vectors a_1 and a_2, by a line in that plane. In the more general case, the plane, spanned by a_1 and a_2, can be replaced by a line that is either closer or further away from the complete markets excess demand of each agent.

This discussion implies that, for each individual, the welfare effect of a monetary union can be decomposed into the gains of changing monetary decision rules and the costs of changing the number of assets in the economy, i.e.,

$$u_u^h - u_0^h = [(\bar{u}_0^h - u_0^h) - (\bar{u}_u^h - u_u^h)] - (\bar{u}_0^h - \bar{u}_u^h).$$

PROPOSITION 2: *In a general equilibrium model with incomplete asset markets, nominal securities, and mean-variance preferences, a monetary union is desirable when the gain from eliminating the excess volatility of nominal variables caused by political shocks exceeds the cost of effectively reducing the number of assets in the economy.*

The implications of this proposition are best understood by considering three special cases: one where a monetary union is Pareto inferior to having independent currencies because monetary policy before the union is already insulated from political shocks; a second one, where a monetary union is Pareto superior because monetary policy before the union is determined by political shocks but after the union it is not; and a third one, that considers fixed exchange rate regimes that fall short of a monetary union.

In the first case, a monetary union is Pareto inferior to a system of independent currencies. Consider a world where there is a representative agent in each country, and agents are heterogeneous across countries. Such a world can be thought of as one, where within each country people experience similar economic shocks or, equivalently, as one where there are good risk-sharing devices within countries but not across national borders.[19] In such a world, independent central banks that follow a constant money supply rule will implement the complete markets allocation. Nominal bonds will be claims on each country's representative agent's income, and will suffice for effectively completing financial markets (the complete markets excess demands will belong to the marketed subspace). If any subset of these countries decides to form a monetary union, the complete markets allocations will no longer be feasible due to the reduction in the number of assets, and the resulting equilibrium will be Pareto inferior to the multicurrency system (independently of the union's monetary policy). This example suggests that, from the perspective of risk sharing, *an optimal currency "area"* is a set of individuals who are able to perfectly share risks among themselves, and use nominal assets to sell their collective risk.[20]

In the second example, a monetary union is Pareto superior. Consider a two-country world where there are no indexed bonds, there is no economic uncertainty, agents want to transfer income intertemporally, and in each country the money supply is very sensitive to political shocks. Consumers will trade these nominal bonds to transfer income intertemporally, and this will make their consumption sensitive to political shocks. A move towards a monetary union with a single monetary authority that follows a constant nominal money supply rule will implement the complete markets allocation that is Pareto superior to the previous equilibrium. In this example, the loss of one asset entails no welfare cost: in the absence of political shocks there is only one state at date 1, and, hence, one asset with a constant real payoff is sufficient to complete the markets.

Third, consider a monetary/exchange rate regime such as the gold standard, the exchange rate mechanism of the ERM, or the currency board implemented in Argentina in 1991. All these arrangements are rules that attempt to

[19] As financial markets and the welfare state are the devices that modern economies have for sharing risks, this situation is consistent with a world where, for some reason, equity-portfolios exhibit a home bias, and where there are no supranational welfare states.

[20] In this respect, it is interesting to observe that the plans for a monetary union in Europe have also motivated proposals for supranational fiscal programs aimed at improving intra-European regional risk sharing (Bean, 1992; Eichengreen, 1993).

limit exchange rate variability, and fall short of an irreversible monetary union (or permanently fixed exchange rates). Under these regimes, the central bank follows a rule that fixes the exchange rate (or allows it to float within a tight band), and agents know that there are circumstances under which the rule can be abandoned (possibly on a temporary basis). Let the set of economic shocks for which the rule will be maintained with probability one be $\mathbf{R} \subset \Sigma$, and let $\mathbf{R}^c$ be its complement. In the states belonging to $\mathbf{R}^c$ it is possible that the rule will be abandoned, and, as the decision to do so is a political one,[21] $\text{var}[(\omega_i/{}_iM)|\sigma]$ will be large. Expressions (6) and (7) imply that credible monetary rules, for which there is a large probability that the rule will not be broken and, therefore, that $\text{var}[(\omega_i/{}_iM)|\sigma]$ will be small, are effective devices for reducing the welfare costs of monetary instability. On the other hand, the extreme circumstances under which policy makers may decide to depart from a preannounced rule (and $\text{var}[(\omega_i/{}_iM)|\sigma]$ may be large) might be precisely the ones where nominal assets are the most useful. For example, countries that are net debtors in their own currency might find it very valuable to repudiate part of their debt through inflation in times of great need (such as wars and major recessions) at the cost of paying higher *ex post* real interest rates in good times. Summing up, the model implies that these hybrid regimes,[22] with both rules and discretion, are a compromise between the benefits of removing all discretionary powers from monetary authorities and the cost of losing the advantages of [illegible]inal contracts as insurance devices.

V. Conclusions

Currency unions are designed with [illegible] purpose of eliminating exchange rate [illegible]ability and insulating monetary policy [illegible] political pressures. At the same time [illegible] change the tools available to share [illegible] among nations. This paper develops a [illegible]eral equilibrium model in which these is[illegible] can be analyzed, and describes an econ[illegible] environment where eliminating exces[illegible] monetary variability improves welfare. [illegible] shown that in an environment with nom[illegible] securities, incomplete markets, and ag[illegible] with quasi-linear quadratic preferences, [illegible] monetary union is desirable when the [illegible] from eliminating the excess volatility [illegible] nominal variables exceed the cost of ch[illegible]ing the asset structure.

The analysis rests on the natural assu[illegible]tions that (monetary) policy depends on [illegible] realization of political events and that in[illegible]ance markets for political risks do not [illegible] Currency unions are viewed as rules th[illegible] eliminate (monetary) political risks by [illegible]ing the link between policy and politics. [illegible] democratic society, however, there are [illegible] reasons for tying policy to political decis[illegible] Cutting this link may result in an allocatio[illegible] resources that is socially undesirable [illegible] though it allocates risks more efficiently. [illegible] interaction between evolving social pre[illegible]ences, public policy, and an efficient alloc[illegible] of risks merits further research.

Appendix

PROPOSITION A1: *Assume that there [illegible] more than two agents in each country ([illegible] 2 for $i = 1, \ldots, I$), asset markets are inc[illegible]plete ($J < S$), and monetary policy, M[illegible] isfies ${}_iM(\sigma, \theta) \neq {}_iM(\sigma, \theta')$ for some σ, [illegible] and i. Then, for a generic set of endow[illegible] there exists a monetary policy $\bar{M}$ (indep[illegible] political shocks) that supports and eq[illegible]rium that Pareto dominates the one sup[illegible] by monetary policy M.*

[21] Examples of political decisions to abandon a monetary rule are ubiquitous in history. Under the classical gold standard, the currencies of the center countries where taken off gold during wars while in the periphery there were also other "critical" circumstances that triggered suspensions of convertibility (Eichengreen, 1992; Michael D. Bordo and Kydland, 1995). In Europe, some governments decided to abandon the ERM after German reunification, while others did not. The Tequila crisis in 1994 (the decision by the Mexican government to abandon its exchange rate regime) put pressure on Argentina to abandon its currency board but the government decided to pay the political costs of sticking to its exchange rate commitment (Cavallo and Cottani, 1997).

[22] For example, ${}_iM_i\omega = {}_jM_j\omega$ in normal times and devalue when ${}_i\omega$ is below a critical unannounced value.

PROOF:

Given any vector $x \in \mathbb{R}^{\Sigma\Theta}$ (except π), let $\bar{x} \in \mathbb{R}^{\Sigma}$ be the projection of x onto $\mathbb{R}^{\Sigma}$, where $\bar{x}(\sigma) = \Sigma_\theta\, \pi(\theta|\sigma)x(\sigma, \theta)$ for all σ. $\bar{x}(\sigma)$ is the expectation of x given σ. For any $\Sigma\Theta \times J$ dimensional matrix $X = (x_1, \dots, x_J)$, let $\bar{X}$ be the $\Sigma \times J$ dimensional matrix $\bar{X} = (\bar{x}_1, \dots, \bar{x}_J)$. For π, $\bar{\pi} = \pi(\sigma) = \Sigma_\theta\, \pi(\sigma, \theta)$. $\Lambda(x)$ is the diagonal matrix derived from the vector x.

A generic set of endowments is an open and dense set of full Lebesgue measure in Ω.

If $\Sigma \geq J$, then there exists a generic set of endowments, Ω^*, such that if $\omega \in \Omega^*$ then $A(p^*)$ has full rank for a given monetary policy (see working paper [1995]).

Simple computations reveal that, in equilibrium,[23]

$$\text{(A1)} \quad y^{h*} = (A(p^*)'\Lambda(\pi)A(p^*))^{-1} \times A(p^*)'\Lambda(\pi)\left(\frac{\omega_1}{H} - \omega_1^h\right),$$

$$\text{(A2)} \quad x^h = \begin{pmatrix} \omega^h(0) - \left(c - \frac{\omega_1}{H}\right)'\Lambda(\pi) \\ \quad \times A(A'\Lambda(\pi)A)^{-1}A'\Lambda(\pi) \\ \quad \times \left(\frac{\omega_1}{H} - \omega_1^h\right) \\ \omega_1^h + A(A'\Lambda(\pi)A)^{-1} \\ \quad \times A'\Lambda(\pi)\left(\frac{\omega_1}{H} - \omega_1^h\right) \end{pmatrix},$$

where the prime $'$ indicates transpose, $\Lambda(\pi)$ is an S-dimensional diagonal matrix of probabilities and $\omega_1 = \Sigma_{h \in \mathbf{H}}^{h}\, \omega_1$ is the world aggregate endowment at date 1. Adding the first-order conditions, $q = (c - x^h)'\Lambda(\pi)A$, over individuals, and using the equilibrium conditions, $\Sigma_h\, x^h = \Sigma_h\, \omega^h$, implies that $q^* = [c - (\omega_1/H)]'\Lambda(\pi)A(p^*)$. This expression for equilibrium asset prices, the first-order conditions, and the budget constraints, in turn, yield $[c - (\omega_1/H)]'\Lambda(\pi)A(p^*) = (c - (\omega^h + Ay^h))'\Lambda(\pi)A$. As, generically, $A(p^*)$ has full rank, $A'\Lambda A$ is invertible, and $y^{h*} = (A'\Lambda A)^{-1}A'\Lambda[\omega_1/H - \omega_1^h]$. Expression (A2) follows from (A1) and the budget constraints.

Expression (A2) and the functional form of the utility function imply that equilibrium utility can be written as:

$$u^h = u(\omega^h) + e^{h\prime}\Lambda A(A'\Lambda A)^{-1}A'\Lambda e^h,$$

where $e^h = (\omega_1/H) - \omega_1^h$. Adding and subtracting $\bar{e}^{h\prime}\bar{\Lambda}\bar{A}(\bar{A}'\bar{\Lambda}\bar{A})^{-1}\bar{A}'\bar{\Lambda}\bar{e}^h$ and noticing that $A'\Lambda e^h = \bar{A}'\bar{\Lambda}\bar{e}^h$ results in

$$\begin{aligned} u^h &= u(\omega^h) + \bar{e}^{h\prime}\bar{\Lambda}\bar{A}(\bar{A}'\bar{\Lambda}\bar{A})^{-1}\bar{A}'\bar{\Lambda}\bar{e}^h \\ &\quad + \bar{e}^{h\prime}\bar{\Lambda}\bar{A}((A'\Lambda A)^{-1} \\ &\qquad - (\bar{A}'\bar{\Lambda}\bar{A})^{-1})\bar{A}'\bar{\Lambda}\bar{e}^h \\ &= \bar{u}^h + \bar{e}^{h\prime}\bar{\Lambda}\bar{A}((A'\Lambda A)^{-1} \\ &\qquad - (\bar{A}'\bar{\Lambda}\bar{A})^{-1})\bar{A}'\bar{\Lambda}\bar{e}^h. \end{aligned}$$

Notice that if asset markets are complete then $u^h = \bar{u}^h$. Consider the term

$$\begin{aligned} &\bar{u}^h - u^h \\ &= \bar{e}^{h\prime}\bar{\Lambda}\bar{A}((\bar{A}'\bar{\Lambda}\bar{A})^{-1} - (A'\Lambda A)^{-1})\bar{A}'\bar{\Lambda}\bar{e}^h \\ &= y^{*h\prime}(A'\Lambda A)(\bar{A}'\bar{\Lambda}\bar{A})^{-1} \\ &\quad \times ((\bar{A}'\bar{\Lambda}\bar{A})^{-1} - (A'\Lambda A)^{-1})y^{*h}. \end{aligned}$$

Observe that $(A'\Lambda A) - (\bar{A}'\bar{\Lambda}\bar{A}) = \left(\begin{smallmatrix} 0 & 0 \\ 0 & \aleph \end{smallmatrix}\right)$, where $\aleph$ is a semipositive definite matrix with generic elements $\Sigma_\sigma\, \pi(\sigma)\mathrm{cov}[1/_ip,\ 1/_jp\,|\sigma]$. It follows that $\bar{u}^h - u^h$ is a positive semidefinite quadratic form and $u^h \geq \bar{u}^h$ for all h. The

[23] Only interior solutions with $x^h \gg 0$ are considered. This requires to assume

$$\begin{pmatrix} \omega^h(0) - \left(c - \frac{\omega_1}{H}\right)'\Lambda(\pi)A(A'\Lambda(\pi)A)^{-1} \\ \quad \times A'\Lambda(\pi)\left(\frac{\omega_1}{H} - \omega^h\right) \\ \quad A(A'\Lambda(\pi)A)^{-1}A'\Lambda(\pi)\left(\frac{\omega_1}{H} - \omega^h\right) \end{pmatrix} \gg 0.$$

columns/rows of $\aleph$ corresponding to currencies with ${}_ip = {}_i\bar{p}$ are zero. Let $\tilde{\aleph}$ be the positive definite matrix that is obtained by deleting the zeros from $\aleph$. Then, the quadratic form can be written as

$$\bar{u}^h - u^h = y^{*h'}\left[(\bar{A}'\bar{\Lambda}\bar{A}) + \begin{pmatrix}0 & 0\\ 0 & \aleph\end{pmatrix}\right] \times (\bar{A}'\bar{\Lambda}\bar{A})^{-1}\begin{pmatrix}0 & 0\\ 0 & \aleph\end{pmatrix} y^{*h}$$

$$= y^{*h'}\left[\begin{pmatrix}0 & 0\\ 0 & \aleph\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & \aleph\end{pmatrix} \times (\bar{A}'\bar{\Lambda}\bar{A})^{-1}\begin{pmatrix}0 & 0\\ 0 & \aleph\end{pmatrix}\right] y^{*h}$$

$$= \tilde{y}^{*h'}(\tilde{\aleph} + \tilde{\aleph}\Delta\tilde{\aleph})\tilde{y}^{*h},$$

where $\tilde{y}^{*h}$ is the equilibrium portfolio containing only securities denominated in currencies with $p \neq \bar{p}$ and Δ is the corresponding positive definite submatrix of $(\bar{A}'\bar{\Lambda}\bar{A})^{-1}$. Adapting an argument from step 3 in the proof of Theorem 1 in Geanakoplos and Mas-Colell (1989), and being careful to keep the price levels in $A(p^*)$ constant when endowments are perturbed, it is possible to show that: if $H_i > 2$ for $i = 1, \ldots, I$, and $J \leq S$, then there exists a generic set of endowments, Ω^{**}, such that for $\omega \in \Omega^{**}$ the matrix $[y^1, \ldots, y^J]$ has full rank, i.e., all the securities are traded (see working paper [1995]). It follows that if $\omega \in \Omega^{**}$, then $\tilde{y}^{*h} \neq 0$ for some h. Therefore, if endowments belong to the generic set $\Omega^* \cap \Omega^{**}$, then $\bar{u}^h - u^h > 0$ for some h.

REFERENCES

Alesina, Alberto and Drazen, Allan. "Why Are Stabilizations Delayed." *American Economic Review*, December 1991, *81*(5), pp. 1170–88.

Balasko, Yves and Cass, David. "The Structure of Financial Equilibrium with Exogenous Yields: The Case of Incomplete Markets." *Econometrica*, January 1989, *57*(1), pp. 135–62.

Bean, Charles. "Economic and Monetary Union in Europe." *Journal of Economic Perspectives*, Fall 1992, *6*(4), pp. 31–52.

Bordo, Michael D. and Kydland, Finn E. "The Gold Standard as a Rule: An Essay in Exploration." *Explorations in Economic History*, October 1995, *32*(4), pp. 423–64.

Calvo, Guillermo A. "Servicing the Public Debt: The Role of Expectations." *American Economic Review*, September 1988, *78*(4), pp. 647–61.

Canzoneri, Matthew B. and Rogers, Carol Ann. "Is the European Community an Optimal Currency Area? Optimal Taxation versus the Cost of Multiple Currencies." *American Economic Review*. June 1990, *80*(3), pp. 419–33.

Cass, David. "On the Number of Equilibrium Allocations with Incomplete Financial Markets." CARESS Working Paper No 85-16, University of Pennsylvania, 1985.

______. "Sunspots and Incomplete Financial Markets: The Leading Example," in George Feiwel, ed., *The economics of imperfect competition and employment: Joan Robinson and beyond*. New York: New York University Press, 1989, pp. 677–93.

Cavallo, Domingo F. and Cottani, Joaquín A. "Argentina's Convertibility Plan and the IMF." *American Economic Review*, May 1997 (*Papers and Proceedings*), *87*(2), pp. 17–22.

Chari, V. V.; Christiano, Laurence J. and Eichenbaum, Martin. "Expectation Traps and Discretion." Working paper, Northwestern University, 1996.

Christiano, Lawrence J. and Eichenbaum, Martin. "Identification and the Liquidity Effect of a Monetary Policy Shock," in Alex Cukierman, Zvi Hercowitz, and Leonardo Leiderman, eds., *Political economy, growth, and business cycles*. Cambridge, MA: MIT Press, 1992, pp. 335–70.

Eichengreen, Barry. *Golden fetters: The gold standard and the Great Depression, 1919–1939*. Oxford: Oxford University Press, 1992.

______. "European Monetary Unification." *Journal of Economic Literature*, September 1993, *31*(3), pp. 1321–37.

Fisher, Irving. *100% money; designed to keep checking banks 100% liquid; to prevent inflation and deflation; largely to cure or prevent depressions; and to wipe out much of the national debt*. New York: Adelphi, 1936.

Frieden, Jeffry A. "Making Commitments: France and Italy in the European Monetary System, 1979–1985," in Barry Eichengreen and Jeffry A. Frieden, eds., *The political economy of European monetary unification*. Boulder, CO: Westview Press, 1994a.

———. "Exchange Rate Politics: Contemporary Lessons from American Politics." *Review of International Political Economy*, Spring 1994b, *1*(1), pp. 81–103.

Friedman, Milton. *A program for monetary stability*. New York: Fordham University Press, 1960.

———. "Should There Be an Independent Monetary Authority?" in Leland B. Yeager, ed., *In search of a monetary constitution*. Cambridge, MA: Harvard University Press, 1962, pp. 219–43.

Friedman, Milton and Schwartz, Anna. *A monetary history of the United States, 1867–1960*. New York: National Bureau of Economic Research, 1963.

Geanakoplos, John and Mas-Colell, Andreu. "Real Indeterminacy with Financial Assets." *Journal of Economic Theory*, February 1989, *47*(1), pp. 22–38.

Helpman, Elhanan. "An Exploration in the Theory of Exchange-Rate Regimes." *Journal of Political Economy*, October 1981, *89*(5), pp. 865–90.

Helpman, Elhanan and Razin, Assaf. "A Comparison of Exchange Rate Regimes in the Presence of Imperfect Capital Markets." *International Economic Review*, June 1982, *23*(2), pp. 365–88.

Kareken, John H., and Wallace, Neil. "On the Indeterminacy of Equilibrium Exchange Rates." *Quarterly Journal of Economics*, May 1981, *96*(2), pp. 207–22.

Kydland, Finn E. and Prescott, Edward C. "Rules Rather than Discretion: The Inconsistency of Optimal Plans." *Journal of Political Economy*, June 1977, *85*(3), pp. 473–91.

Lucas, Robert E., Jr. "Interest Rates and Currency Prices in a Two-Country World." *Journal of Monetary Economics*, November 1982, *10*(3), pp. 335–59.

Magill, Michael and Quinzii, Martine. "Real Effects of Money in General Equilibrium." *Journal of Mathematical Economics*, 1992, *21*(4), pp. 301–42.

———. *Theory of incomplete markets*. Cambridge, MA: MIT Press, 1996.

Mundell, Robert A. "A Theory of Optimum Currency Areas." *American Economic Review*, September 1961, *51*(4), pp. 657–65.

———. "Uncommon Arguments for Common Currencies," in Harry Johnson and Alexander Swoboda, eds., *The economics of common currencies*. Cambridge, MA: Harvard University Press, 1973, pp. 114–32.

Mussa, Michael. "Political and Institutional Commitment to a Common Currency." *American Economic Review*, May 1997 (*Papers and Proceedings*), *87*(2), pp. 217–20.

Neumeyer, Pablo Andrés. "Currencies and the Allocation of Risk: The Welfare Effects of a Monetary Union." DELTA Working Paper No. 95-27, September 1995.

Persson, Torsten and Svensson, Lars E. O. "Exchange Rate Variability and Asset Trade." *Journal of Monetary Economics*, May 1989, *23*(3), pp. 485–509.

Polemarchakis, Heracles. "Portfolio Choice, Exchange Rates and Indeterminacy." *Journal of Economic Theory*, December 1988, *46*(2), pp. 414–21.

Siconolfi, Paolo. "Sunspot Equilibria and Incomplete Financial Markets." *Journal of Mathematical Economics*, 1991, *20*(3), pp. 327–39.

Svensson, Lars E. O. "Trade in Nominal Assets: Monetary Policy, and Price Level and Exchange Risk." *Journal of International Economics*, February 1989, *26*(1/2), pp. 1–28.

The Ultimate Externality

By JON D. HARFORD*

A unified formal theoretical framework for addressing externalities and population issues seems to have heretofore been lacking. Kerstin L. Kiessling and Hans Landberg (1994 p. xviii) clearly state that population growth has an external effect, but refer to no formal model. Partha Dasgupta (1993 pp. 350–51) argues informally (only) that childbearing creates an external effect when common property resources are not properly priced. In a paper with results in the same spirit as those found here, John O. S. Kennedy (1995) analyzes the impact of immigration in the presence of negative externalities and finds that the original population can be harmed by immigration even with externality taxation.

The present paper addresses the issues of childbearing choices and externalities in a two-generation model of a competitive economy with a pollution externality created by the aggregate consumption of one of two goods. For concreteness, the externality will be interpreted as pollution, but other interpretations are possible. In fact, an interpretation based upon a public good externality, such as might result from educational activity, is entirely compatible with the mathematical model.

It is assumed that the utility of each member of the final generation is a function of her own consumption and the level of pollution. Each member of the initial generation experiences utility based on her own consumption and pollution exposure, but each also derives utility from the utility of her children. The combination of a narrow intergenerational concern and a public bad externality leads to the result that Pareto efficiency requires not only Pigovian pollution taxes, but a tax per child the size of which approximately equals the present value of the pollution taxes each descendant will pay. Intuitively, this result follows from the fact that each member of the initial generation should pay a Pigovian tax for each externality her actions create. In this case, since the parent chose to have a child, she should be charged for the externality created by her child.

The present results can be better understood by comparing them with those of William Schulze and Ralph C. d'Arge (1974) and Daniel F. Spulber (1985) regarding the Pigovian taxation of pollution from firms. Such taxation gives each existing firm the proper incentive to limit pollution, and efficiently limits the number of firms by raising the equilibrium average cost via the added tax burden. These authors point out that tax revenues must be rebated in lump-sum fashion because rebates to firms would inappropriately offset the entry-limiting effects of pollution taxes. In the present context, all individuals in the final period receive a lump-sum rebate of taxes equal in size to equilibrium taxes paid within that period. This implies that an additional child creates no fiscal externality in her generation to offset the technological externality that she creates. Thus, in the absence of an appropriate childbearing tax, a member of the initial generation has an inadequate incentive to limit the "entry" of her children into the final generation.[1]

I. The Two-Period Model

There are two generations: the generation of the parents (generation 0) and the generation

* Department of Economics, Cleveland State University, Cleveland, OH 44115. The author wishes to thank John O. S. Kennedy and his colleagues Myong Chang and David Yerger for feedback regarding the ideas in this paper. The author also thanks three referees for many detailed comments that helped shape the final version of this paper and eliminate errors. The author takes responsibility for any remaining errors.

[1] Another perspective on the need for both pollution and childbearing taxes has been offered [by a co-editor]. Total pollution in the final generation is a product of the population and per capita consumption of the polluting good. To control both population and per capita consumption one needs two instruments.

of the children (generation 1). Each generation lives one period and the periods do not overlap. The population of the parents' generation is N_0 and the population of the children's generation is N_1. Each member i of generation 0 produces n_{0i} children in an asexual manner so that each child has only one parent. Each member of the initial generation has the same income and tastes, and each produces children that are like all other children in tastes and equilibrium income. It is assumed that all optimizing problems discussed result in interior solutions. All variables, including the number of children, will be treated as continuous.

The concave utility function of each member j of the final generation can be written $U_{1j}[c_{1j}, x_{1j}, A_1]$, where "$c$" and "$x$" refer to consumption goods that are produced in competitive markets. "A_1" is interpreted as a public bad and will be called pollution. Specifically,

$$A_1 = \sum_{m=1}^{N_1} x_{1m}.$$

Thus, A_1 is aggregate consumption of good x in the final generation.

It is assumed that j will maximize her utility subject to a budget constraint of the form

$$(1) \qquad b_{1*} + I_{1j} = c_{1j} + (q_1 + t_1)x_{1j},$$

where q_1 is the competitive supply price of good x_1 in terms of the numeraire good c_1; t_1 is a constant unit tax on good x_1; b_1 is a lump-sum rebate which will equal the equilibrium per capita amount of taxes paid ($b_{1*} = t_1 x_{1*}$); and I_{1j} is a level of income that is exogenous from the viewpoint of person j, but partially determined by the bequest of capital left to j by her parent i. The presence of tax-related terms in (1) and other equations is in anticipation of the policy role they will play.

The income of person j is $I_{1j} = k_{0i}(1 + r_1) + w_1$, where w_1 is the competitive wage for the one unit of labor each person is endowed with, k_{0i} is the bequest of capital received by each child of i, and r_1 is the competitive rate of return on capital in generation 1's time. Furthermore, $k_{0i} = (K_{0i}/n_{0i})$, where K_{0i} is the total bequest left by i to her n_{0i} children.

Given this budget constraint, a member j of generation 1 would maximize utility by satisfying the marginal condition

$$(2) \quad (\partial U_{1j}/\partial x_{1j})/(\partial U_{1j}/c_{1j}) = q_1 + t_1 - (\partial U_{1j}/\partial A_1)/(\partial U_{1j}/c_{1j}).$$

Equation (2) indicates that the marginal rate of substitution of c_1 for x_1 by j will equal the full price of x_1 in terms of c_1, which includes the "nuisance" term on the far right. This term reflects how many units of good c that j would give up to reduce a unit of her pollution exposure that she influences by her consumption of good x. Equations (1) and (2) define the choices of x_{1j} and c_{1j} as functions of $(I_{1j} + b_{1*})$ and t_1.

The utility function of a member i of the initial generation is assumed to be

$$(3) \qquad V_{0i} = U_{0i}[c_{0i}, x_{0i}, A_0] + \beta n_{0i} U_{1j}[c_{1j}, x_{1j}, A_1],$$

where U_{0i} is a concave function of initial period consumption and pollution, U_{1j} is the utility of each of the children of person i of the initial generation, and

$$A_0 = \sum_{s=1}^{N_0} x_{0s}.$$

Person i weights the sum of utilities experienced by her children by the generational discount factor β, where $0 < \beta < 1$. The definition of V_{0i} roughly captures the idea that a parent cares more about her children than about unrelated individuals. Except for the externality, the assumptions on V_{0i} are similar to those of Gary S. Becker and Robert J. Barro (1988).[2]

[2] Instead of the constant discount factor β, Becker and Barro use the form $a[n] = \alpha n^{-\varepsilon}$ with $0 < \alpha < 1$ and $0 < \varepsilon < 1$. Becker and Barro's more general form implies the per-child generational discount factor shrinks as the number of children increases but that $a[n]n$ increases in n. The basic nature of the results of this paper would remain unchanged for any discount factor function with n as its sole argument as long as it produced an interior solution to the optimizing problem.

The income I_{0i} of a member i of generation 0 is considered a parameter that has the value I_0 for all i. The budget constraint for i will be written

$$(4)\quad b_{0*} + I_{0i} = c_{0i} + (q_0 + t_0)x_{0i} + (p + T)n_{0i} + K_{0i},$$

where q_0 is the competitive price of good x_0 in terms of numeraire good c_0; p is the constant cost per child; K_{0i} is the total cost of all bequests to her children; t_0 is a constant unit tax on good x_0; and T is a constant tax per child. Finally, $b_{0*} = t_0 x_{0*} + Tn_{0*}$, where x_{0*} and n_{0*} refer to the common equilibrium values for x_0 and n_0, is the per capita lump-sum rebate of tax collections.

Maximization of V_{0i} by the choices of c_{0i}, x_{0i}, n_{0i}, and K_{0i} subject to (1) and (4) yields the following marginal conditions:

$$(5a)\quad (\partial U_{0i}/\partial x_{0i})/(\partial U_{0i}/\partial c_{0i}) = q_0 + t_0 - (\partial U_{0i}/\partial A_0)/(\partial U_{0i}/\partial c_{0i}),$$

$$(5b)\quad (\partial U_{1j}/\partial c_{1j})/(\partial U_{0i}/\partial c_{0i}) = 1/\{\beta(1 + r_1)\},$$

$$\begin{aligned}(5c)\quad p + T &= \{\beta U_{1j}/(\partial U_{0i}/\partial c_{0i})\} - \{(\partial U_{1j}/\partial c_{1j})/(\partial U_{0i}/\partial c_{0i})\} \times (1 + r_1)\beta k_{0i} \\ &\quad - \beta n_{0i}\{(\partial U_{1j}/\partial A_1) \div (\partial U_{0i}/\partial c_{0i})\}x_{1j} \\ &= \{\beta U_{1j}/(\partial U_{0i}/\partial c_{0i})\} - k_{0i} - \beta n_{0i}\{(\partial U_{1j}/\partial A_1) \div (\partial U_{0i}/\partial c_{0i})\}x_{1j}.\end{aligned}$$

Equation (5a) has an obvious interpretation analogous to (2). Equation (5b) indicates that (at the margin) the number of units of her consumption she will give up to increase each of her children's consumption by one unit is inversely related to the product of the generational discount factor and the gross rate of return on capital. Equality (5b) was used to obtain the final version of the right-hand expression in (5c). A rearrangement of (5c) indicates that the utility-maximizing number of children will be such that the full private cost of a child $(p + T + k_{0i})$ times the marginal utility of income to the parent will equal the increase in utility derived from having one more child endowed with the same resources as the others. The increase in utility from having one more child is slightly reduced by the far-right term in (5c), which reflects the impact that one more child will have on her siblings via the increased pollution she creates.

II. Pareto-Efficient Resource Allocation

It is assumed that for each generation there is a separate all-knowing, costlessly acting planner who has the goal of achieving Pareto efficiency from the viewpoint of members of her generation. The rationale for assuming a separate planner for each generation is that this more accurately reflects the realities of policy making.[3] Each planner takes the other's actions as given. In generation 1, Pareto efficiency will be produced by maximizing the utility of the "representative individual," whose choices are notationally reflected in the dropping of individual subscripts. In effect, the

[3] Two virtually indisputable assumptions lead one to the conclusion that the only welfare functions the policy implications of which have any hope of becoming reality (except by accident) are those which are a function of the preferences of members of the current generation. The first assumption is that individuals act according to their preferences. (This is really just a definition of preferences.) The second assumption is that only the living can directly affect policy choices made in their generation. In other words, the dead and the unborn do not vote. From a biological viewpoint, this author would presume that we have altruistic preferences towards our children precisely because they do not have the power to create and maintain themselves until they become members of the next parental generation.

planner chooses the common values of c_1 and x_1 subject to the per capita income constraint $(1 + r_1)k_0 + w_1 = c_1 + q_1x_1$. The planner recognizes the impact of x_1 on the value of $A_1 = N_1x_1$. Maximizing $U_1[c_1, x_1, A_1]$ subject to the income constraint yields the efficiency condition

$$(6) \quad (\partial U_1 / \partial x_1)/(\partial U_1 / c_1) = q_1 - N_1(\partial U_1 / \partial A_1)/(\partial U_1 / \partial c_1),$$

where q_1 is interpreted as both supply price and equilibrium rate of product transformation. The marginal rate of substitution between x_1 and c_1 displayed in (2) will equal that required for Pareto efficiency as indicated in (6) if

$$(7) \quad t_1 = -(N_1 - 1)(\partial U_1 / \partial A_1)/(\partial U_1 / \partial c_1)$$

evaluated at the efficient allocation. The right-hand side represents the negative of the sum of marginal rates of substitution of the numeraire good for pollution for the $(N_1 - 1)$ "other" members of generation 1. This is the traditional Pigovian tax.

Pareto efficiency for the initial generation will be achieved if the representative individual achieves a maximum level of V_0 with respect to the choices of the common values of c_0, x_0, n_0, and K_0 subject to the per capita income constraints on generation 1 and on generation 0, with the latter being $I_0 = c_0 + q_0x_0 + pn_0 + K_0$. Thus, the initial-period planner will recognize the impact of x_0 on $A_0 = N_0x_0$ and the impact of n_0 on A_1 through the relationship $A_1 = N_1X_1 = N_0n_0x_1$. Assuming that (6) is satisfied, the necessary conditions for Pareto efficiency for the initial generation are

$$(8a) \quad (\partial U_0 / \partial x_0)/(\partial U_0 / c_0) = q_0 - N_0(\partial U_0 / \partial A_0)/(\partial U_0 / c_0),$$

$$(8b) \quad (\partial U_1 / \partial c_1)/(\partial U_0 / \partial c_0) = 1/\{\beta(1 + r_1)\},$$

$$(8c) \quad p = \{\beta U_1 / (\partial U_0 / \partial c_0)\} - k_0 + \beta n_0 N_0 x_1 \{(\partial U_1 / \partial A_1) \div (\partial U_0 / \partial c_0)\}.$$

Comparing equation (8a) to (5a), it follows that individual trade-offs can be made consistent with efficient trade-offs between the goods c and x in period 0 by setting $t_0 = -(N_0 - 1)(\partial U_0/\partial A_0)/(\partial U_0/c_0)$.

Equation (8b) is identical to (5b) except for the individual subscript, indicating that the unmodified individual trade-off between current consumption and total bequest is consistent with efficiency, as long as t_1 is determined by (7). This result is more interesting than the reader might suspect. Martin L. Weitzman (1994) used a bare-bones model of a dynamic economy to argue that pollution externalities imply a social discount rate that is below the unadjusted rate of return on capital due to what he called "environmental drag." The present model shows no such divergence between the unadjusted rate of return on capital and the social discount rate. However, Harford (1996) has shown that results similar in spirit to Weitzman's, but significantly different in detail, occur when the present type of model is modified so that the production of the capital good is polluting.[4] Thus, the present results indicate that the social discount rate will not differ from the unadjusted rate of return when pollution results from producing a noncapital good, while Harford's result confirms that "environmental drag" will modify the discount rate when production of the capital good is polluting.

In equation (8c) I have used (8b) to reduce one term to k_0. A comparison of (5c) with (8c) indicates that individual utility

[4] Weitzman's model has only one good and no explicit utility function. His expression for the social discount rate is algebraically and conceptually more complicated than Harford's (1996). The reader is referred to equations (17) and (18) in Weitzman (1994 p. 205). In the Harford model the social discount rate becomes $(1 - t)r$, where r is the unadjusted rate of return on capital and t is the pollution tax rate as applied to the appropriate generation.

maximization can be consistent with Pareto efficiency only if

$$(9)\quad T = -\beta n_0(N_0 - 1)x_{1*} \times \{(\partial U_1/\partial A_1)/(\partial U_0/\partial c_0)\} = \{t_1 x_{1*}/(1 + r_1)\} \times \{(N_1 - n_0)/(N_1 - 1)\},$$

where the last expression on the right results from using (8b).

As (9) indicates, the childbearing tax on the parent should equal in value the present value of the pollution taxes that a child would pay in her lifetime times a factor that will equal one if parents each have one child. The factor is less than one if each parent has more than one child since the parent internalizes the effect on siblings of pollution from each child.

The first equality in equation (9) has an intuitive interpretation. If another unit of pollution is created in the final generation, each parent in the initial generation will suffer a loss equal to $-\beta n_0\{(\partial U_1/\partial A_1)/(\partial U_0/\partial c_0)\}$ when evaluated in terms of the numeraire good. Therefore, a parent giving birth to another child producing additional pollution x_{1*} imposes a loss on $(N_0 - 1)$ other parents equal to $-\beta n_0(N_0 - 1)x_{1*}\{(\partial U_1/\partial A_1)/(\partial U_0/\partial c_0)\}$. The childbearing tax thus prices the external cost a child imposes on other parents via the external cost a child imposes on their children.[5]

The model can be readily extended to many generations. In such a model the extended utility function of a member of generation 0 would contain a term for each future generation g of the form $(\beta^g n_0 n_1 \cdots n_{g-1} U_g)$, where the pollution argument of U_g will have the form $A_g = N_0 n_0 n_1 \cdots n_{g-1} x_g$. It has been shown by Harford (1996) that the total Pigovian childbearing taxes on person i of generation 0 would equal the discounted present value of the pollution taxes paid by all descendants of i in all generations.

III. Concluding Remarks

With a mixture of positive and negative externalities in a two-generation model it is clear that the childbearing tax would approximately equal the per capita *net* present value of Pigovian taxes and subsidies paid by one's descendants. Among economists and others, there is disagreement about whether population growth is a "problem," which in present terms can be interpreted as a disagreement about the sign of the net value of externalities caused by one more child.[6] This disagreement points to the existence of costly and imperfect information, the same problems that arise in the implementation of virtually any policies intended to affect childbearing or other externalities.[7] While the present paper ignores these kinds of potential complications, it has the virtue of allowing both childbearing and externality-generating consumption choices to be endogenous in a model where the disproportionate concern of parents for their own children is recognized.

REFERENCES

Becker, Gary S. and Barro, Robert J. "A Reformulation of the Economic Theory of Fertility." *Quarterly Journal of Economics*, February 1988, *103*(1), pp. 1–25.

Dasgupta, Partha. *An inquiry into well-being and destitution*. Oxford: Clarendon Press, 1993.

Harford, Jon D. "Pollution, Child-Bearing Externalities, and the Social Discount Rate." Presented at the Eastern Economic Association Meetings, Boston, MA, March 1996.

Kennedy, John O. S. "Changes in Optimal Pollution Taxes as Population Increases."

[5] Leslie A. Whittington et al. (1990) found that the personal exemption for dependents has had a positive and significant effect on the national birth rate of the United States. This suggests that a childbearing tax would have a significant impact on such choices.

[6] Julian L. Simon (1981) (whose book title inspired the title of this paper) argues that additional population increases the rate of technological improvement. He apparently feels that positive population externalities dominate, and thus his views contrast substantially with those of authors such as Dasgupta.

[7] Spulber (1989) analyzes various regulatory problems involving issues of asymmetric information and costly monitoring that would be relevant to population issues.

Journal of Environmental Economics and Management, January 1995, *28*(1), pp. 19–23.

Kiessling, Kerstin L. and Landberg, Hans, eds. *Population, economic development, and the environment*. Oxford: Oxford University Press, 1994.

Schulze, William and d'Arge, Ralph C. "The Coase Proposition, Information Constraints, and Long-Run Equilibrium." *American Economic Review*, September 1974, *64*(4), pp. 763–72.

Simon, Julian L. *The ultimate resource*. Princeton, NJ: Princeton University Press, 1981.

Spulber, Daniel F. "Effluent Regulation and Long Run Optimality." *Journal of Environmental Economics and Management*, June 1985, *12*(2), pp. 103–16.

———. *Regulation and markets*. Cambridge, MA: MIT Press, 1989.

Weitzman, Martin L. "On the 'Environmental' Discount Rate." *Journal of Environmental Economics and Management*, March 1994, *26*(3), pp. 200–09.

Whittington, Leslie A.; Alm, James and Peters, H. Elizabeth. "Fertility and Personal Exemption: Implicit Pronatalist Policy in the United States." *American Economic Review*, June 1990, *80*(3), pp. 545–56.

The Impact of Educational Standards on the Level and Distribution of Earnings

By JULIAN R. BETTS*

In recent years, the role of standards in improving the quality of public education has received considerable attention from economists. Recent examples include Suk Kang (1985), William Becker and Sherwin Rosen (1990), and Robert M. Costrell (1994).[1] These papers are important in the sense that they view education as a principal-agent problem. This contrasts with the main paradigm for empirical work on school quality, which attempts to estimate a "production function," in which purchased inputs such as teachers and length of the school year are assumed to add value to the "intermediate input," students.[2] The theoretical literature on standards points out that academic achievement may have as much to do with incentives as with measures of school resources such as spending per pupil.

An implication of the existing theoretical models of educational standards is that an egalitarian planner may prefer lower standards than an income-maximizing planner, because an increase in standards will aggravate inequality. This point is made most clearly and thoroughly by Costrell (1994) who, unlike Kang (1985) and Becker and Rosen (1990), models earnings as an endogenous function of educational achievement. An increase in educational standards raises wages for those workers who choose to meet the standard while leaving the wages of those who fail to meet the standard unchanged. Costrell (1994, p. 960) concludes that "As egalitarian policy makers lower standards to raise the graduation rate, they reduce GDP, providing an important example of the classic trade-off between the size of the pie and its equal division."

The goal of this paper is to present a different finding: an egalitarian policy maker might prefer *higher* standards than would a policy maker whose goal was to maximize the sum of earnings. The result is based on the observation that if workers are differentiated with respect to ability, an increase in educational standards will increase the earnings of both the most-able and the *least*-able workers. The only workers whose earnings fall are those workers who after the increase fail to continue meeting the standard.

Section I develops the model. The closest parallel in the literature is Costrell (1994). The key distinction between the present model and Costrell's is that the present model assumes that students differ in ability. In the Costrell model, all students are assumed to be equally productive if they exert zero effort, but students do differ in preferences. The model developed here is similar to the models of Kang (1985) and Becker and Rosen (1990) in that it assumes heterogeneous abilities. But instead of assuming that the "rewards" attached to meeting the educational standard are fixed, as do these two papers, the present paper, like that of Costrell, allows the resulting wage to be an endogenous function of the productivities of workers who select into each group.

* Department of Economics, University of California-San Diego, La Jolla, CA 92093. This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant #RED-9452861. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies. I also thank UCSD for research support for initial phases of this work. I thank Robert Costrell, Vincent Crawford, and Joel Sobel for helpful suggestions.

[1] A related contribution by Andrew Weiss (1983) carefully documents that education can act both as a sorting mechanism and a source of human capital if students are passed or failed based on a test of achievement. The paper by Weiss does not focus on the impact of a change in educational standards per se. For related work on how firms may use apprenticeships or other "tests" to identify the best workers, see J. Luis Guasch and Weiss (1980, 1982).

[2] See Eric A. Hanushek (1986, 1989, 1991, 1996) for excellent reviews of the literature on education production functions. For recent reviews on the relation between school resources and earnings, see Betts (1996) and David Card and Alan B. Krueger (1996).

This is a useful extension since it uses a standard utility function to explain why students would care about their performance in school, without recourse to ad hoc assumptions about why meeting the educational standard affects utility.

I. The Model

Consider a hybrid human capital/signaling model in which attendance at school adds to productivity. Firms can observe the standard which each student meets. In particular, they can identify those who graduate from high school and those who do not. This information signals to firms the expected productivity of each worker.

Each student has an initial level of achievement (or ability) of a, where a is a random variable distributed on $[\underline{a}, \bar{a}]$ with probability density and cumulative distribution functions $f(a)$ and $F(a)$, respectively. Student i maximizes utility, which depends on leisure (L_i) and lifetime earnings (w_i):

$$(1) \quad U = U(L_i, w_i) \quad \text{subject to } L_i \in [0, \bar{L}].$$

Educational achievement, π, is defined by an educational production function g, which depends on both ability and effort. The value marginal product of a worker of ability a_i who has chosen a level L_i of leisure is directly proportional to educational achievement, with α denoting the value marginal product of one unit of educational achievement:

$$(2) \quad VMP_i = \alpha\pi_i = \alpha g(\bar{L} - L_i, a_i), \quad \text{s.t. } g_1 > 0,\ g_2 > 0,\ g_{11} \le 0,\ g_{22} \le 0,\ g_{12} \ge 0 \text{ and } \alpha > 0.$$

Note that the first argument of g, $\bar{L} - L_i$, equals student effort. The school sets a standard for the minimum level of achievement necessary to pass, which is denoted π_s.

Firms cannot observe the productivity of workers directly. But firms do observe whether the student has met the standard imposed by the school, and thus they infer his expected productivity. Perfect competition in the labor market ensures that workers are paid their expected product conditional on whether they have met the standard. Workers belong either to group 1, which consists of students who have met or exceed the standard π_s, or group 2, which consists of workers who have not met the standard. Wages will be constant within each group due to the unobservability of ability. Conditional on reaching the achievement level needed to belong to group j, $j = 1, 2$, the worker optimizes by choosing the maximum level of leisure possible.

RESULT 1: *If some students choose to reach the standard and others do not, then there exists an ability level a^* such that workers choose group (1/2) as $a \gtrless a^*$. For workers in group 2, who do not meet the standard, $L_i = \bar{L}$. For $\pi_s < g(0, \bar{a})$, the educational achievement of students is given by the following equation:*

$$(3) \quad \pi_i = \begin{cases} g(0, a_i) < \pi_s, & \text{if } a_i < a^* \\ \pi_s & \text{if } a_i \in [a^*, a^{**}] \\ g(0, a_i) > \pi_s & \text{if } a_i \in (a^{**}, \bar{a}], \end{cases}$$

where a^{**} is implicitly defined by $g(0, a^{**}) \equiv \pi_s$, and a^* is the ability level at which a worker is indifferent between belonging to groups 1 or 2.[3]

PROOF:

For all workers in group 2, wages and leisure are identical at values of $\alpha E(\pi | L = \bar{L})$ and $\bar{L}$. Utility is thus independent of ability in group 2. But utility is nonstrictly increasing in ability if a worker chooses group 1, the members of which must achieve or surpass the standard π_s:

[3] If standards are high enough, so that $\pi_s \ge g(0, \bar{a})$, there will be no students who exceed the standard. In this case, the equations in the model can be simplified by setting $a^{**} = \bar{a}$. All of the results in the model continue to hold in this simplified case.

$$(4)\qquad \left.\frac{dU}{da}\right|_{\pi\geq\pi_s} = \frac{dU}{dL}\left(\left.\frac{dL}{da}\right|_{\pi\geq\pi_s}\right)$$

$$\begin{cases} >0 & \text{for } a\in[\underline{a},a^{**}) \\ =0 & \text{for } a\geq a^{**}. \end{cases}$$

Thus if workers of ability a^* are indifferent between the two groups, all those of (lower/higher) ability will strictly prefer group (2/1). The remaining statements follow directly.

The equations which follow assume that the standard is nontrivial, in that at least some of the students who meet the standard must exert positive effort to do so: $a^* < a^{**}$.

The expected productivity, and hence the wage, for workers in each group can now be calculated:

$$(5)\qquad w^1 = \alpha E(\pi_i \mid \pi_i \geq \pi_s)$$

$$= \frac{\int_{a^*}^{a^{**}} \alpha\pi_s f(a)\,da + \int_{a^{**}}^{\bar{a}} \alpha g(0,a)f(a)\,da}{1-F(a^*)}$$

and for workers who do not meet the standard,

$$(6)\qquad w^2 = \alpha E(\pi_i \mid \pi_i < \pi_s)$$

$$= \frac{\alpha\int_{\underline{a}}^{a^*} g(0,a)f(a)\,da}{F(a^*)},$$

where all workers optimize by setting $L_i = \bar{L}$.

What happens when schools raise the standard π_s? For students who continue to meet or exceed the standard, their average level of achievement is likely to rise, and hence so will their wage:

$$(7)\qquad \left.\left(\frac{dw^1}{d\pi_s}\right)\right|_{a>a^*} = \frac{\alpha(F(a^{**})-F(a^*))}{1-F(a^*)} + \frac{dw^1}{da^*}\frac{da^*}{d\pi_s}.$$

The first term is positive. It is straightforward to show that $dw^1/da^* \geq 0$. The sign of $da^*/d\pi_s$ is likely to be positive, as will be discussed below.

For students of lower ability, the effect of higher standards depends crucially on the sign of $da^*/d\pi_s$. As a^* increases, the average ability of workers in the bottom group rises, thereby increasing the wage paid to this group. More formally, differentiate (6) with respect to π_s:

$$(8)\qquad \left.\left(\frac{dw^2}{d\pi_s}\right)\right|_{a<a^*} = \frac{dw^2}{da^*}\cdot\frac{da^*}{d\pi_s} = \alpha\frac{f(a^*)}{F(a^*)} \times\left\{ g(0,a^*) - \frac{\int_{\underline{a}}^{a^*} g(0,a)f(a)\,da}{F(a^*)}\right\} \times\frac{da^*}{d\pi_s}.$$

The term in braces is the gap between the productivity of a worker of ability a^* and the average productivity of workers of ability below a^*, subject to all such workers choosing $L = \bar{L}$. Since this term is positive (i.e., $dw^2/da^* > 0$), the wage paid those in the lower group rises or falls with a rise in school standards as $(da^*/d\pi_s) \gtrless 0$.

I will make the reasonable assumption that after an increase in educational standards fewer students will meet the standard. In other words, $da^*/d\pi_s > 0$. This will in fact be shown to hold for any stable equilibrium. First, let a_m denote the ability level of the marginal worker (in equilibrium or out of equilibrium). Denote the marginal worker's utility gain from meeting the standard versus not meeting the standard by

$$(9)\qquad \Delta(a_m,\pi_s) \equiv U[L^*(a_m,\pi_s), w^1(a_m,\pi_s)] - U[\bar{L}, w^2(a_m)],$$

where $L^*(a_m, \pi_s)$ is implicitly defined by $\pi_s = g(\bar{L} - L^*, a_m)$. In equilibrium,

[4] I thank Robert Costrell for suggesting this argument.

marginal worker is indifferent, so $\Delta(a^*, \pi_s) = 0$. The comparative static derivative is $da^*/d\pi_s = -(\partial\Delta/\partial\pi_s)/(\partial\Delta/\partial a_m)$ at $a_m = a^*$.

One can unambiguously sign $\partial\Delta/\partial\pi_s < 0$. Figure 1 illustrates, showing the indifference curve and production possibilities frontier for a person of ability a^*, who is originally indifferent between the pairs $(\bar{L}, w^2)$ and $(L^*(a^*, \pi_s), w^1)$.[5] Consider the effect of an increase in the standard and the accompanying increase in the wage paid to those meeting the standard. A person of ability a^* will lower his utility (to the dotted indifference curve) if he chooses to meet the new standard. (A leftward move along the production possibilities frontier moves him onto a lower indifference curve.)

Stability is now invoked to obtain the sign of $\partial\Delta/\partial a_m$. Suppose that the margin a_m varies instantaneously in response to the utility gain from meeting the standard:

$$(10)\quad da_m/dt = \theta[\Delta(a_m, \pi_s)],\ \theta[0] = 0,$$

$$\theta'[0] < 0.$$

The local stability condition is that da_m/dt must decline as a_m rises. Applying the chain rule, the condition is $\theta'[0](\partial\Delta/\partial a_m) < 0$. It immediately follows that at any stable equilibrium, $\partial\Delta/\partial a_m > 0$, and thus $da^*/d\pi_s > 0$.

Although the educational achievement of individual workers in group 2 remains constant, the average productivity of workers in this group rises with a^*, as can be seen from (8). *Thus the wage paid to workers who do not meet the standard in fact rises.*[6] To summarize:

RESULT 2: *The impact of a rise in π_s on a^* is in theory ambiguous, but in any stable equilibrium, $da^*/d\pi_s > 0$. In this case, if the educational standard is increased, educational achievement rises for those in the top group, although if the most-able students can exceed the standard with zero effort, their achievement will not change after a change in the standard. After the increase in the standard, achievement falls for the infinitesimal mass of workers at the margin, and remains constant for workers who before and after the increase in standards choose group 2. The most important conclusion is that a marginal increase in educational standards increases the wages of the workers of highest and lowest ability. The only workers whose wages drop are the workers on the margin who no longer meet the standard. Although of infinitesimal mass, these marginal workers suffer discrete wage drops, while all other workers experience marginal wage gains.*

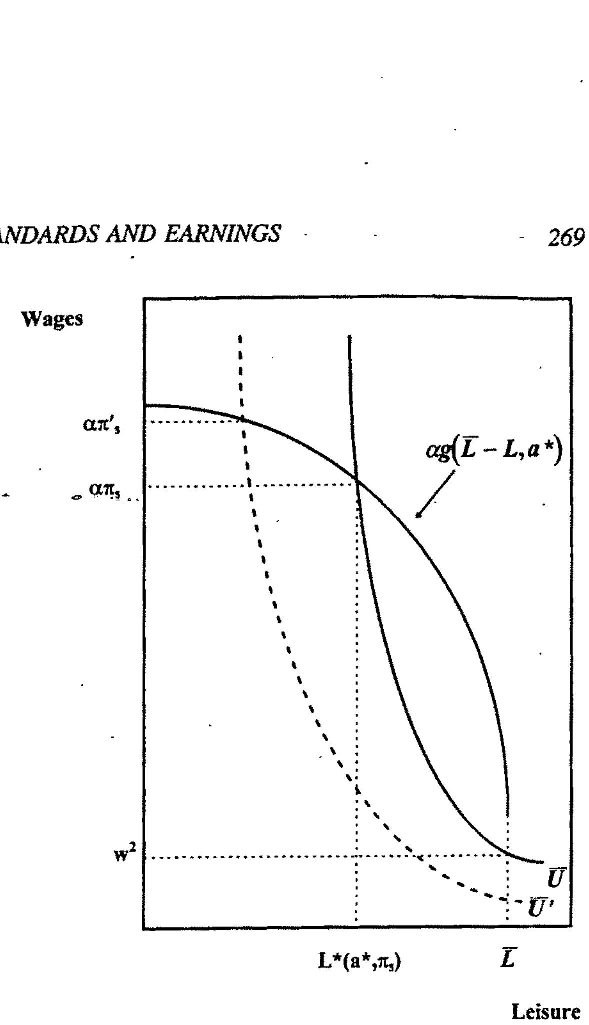


FIGURE 1. OPTIMAL CHOICE OF WAGE AND LEISURE FOR WORKER INDIFFERENT BETWEEN MEETING AND NOT MEETING THE STANDARD

Note: The figure assumes that no worker exceeds the standard, so that $w^1 = \alpha\pi_s$.

II. Implications

A. *Solving the Planner's Problem*

The main result of the paper is that a social planner who values equity may set *higher*

[5] In order to make the relationship between the standard and workers' choice clear, in the figure I assume that no workers exceed the standard, so that $w^1 = \alpha\pi_s$. This is not necessary for the stated results to be true.

[6] An unpublished version of Costrell (1994) makes a similar point. See Costrell (1993 footnote 9).

standards than a social planner whose goal is simply to maximize the sum of earnings. This differs from the findings of Costrell (1994). It derives from the earlier result that in a model with heterogeneous but imperfectly observable abilities, the wages of both the most able and the *least* able will rise after an increase in the educational standard. Following Costrell, assume that the social planner does not value the leisure of students, and instead seeks to maximize some positive function of students' earnings after they leave school.

RESULT 3: *Consider an egalitarian planner who maximizes the sum of a concave function of workers' incomes:* $V^E = F(a^*)h[w^2(a^*)] + (1 - F(a^*))h[w^1(a^*, \pi_s)]$, *where* $dh(y)/dy > 0$ *and* $d^2h/dy^2 < 0$. *This planner may choose a standard higher or lower than would the income-maximizing planner.*

PROOF:

See the Appendix.

Thus in a model with heterogeneous abilities, we cannot assume that concerns about inequality should oblige policy makers to set lower standards. A more egalitarian planner may prefer a higher standard because it leads to higher average ability in the bottom group of students, and thus higher wages for the less-able students, a benefit which is more highly valued by the egalitarian policy maker. As another way of seeing this, note that if a Rawlsian planner were given a choice between two educational standards, both of which entailed some students meeting the standard, the planner would always prefer the higher standard because it is closer to the pooling equilibrium which will maximize the earnings of the least-able worker. The model of Costrell does not come to the same conclusion because in that model all workers who fail to meet the educational standard are equally productive in the labor market, and so the wage paid to this group of workers is independent of the group's membership.[7]

B. *An Example*

Consider a simple example of the above model which illustrates that it is indeed possible that a more egalitarian planner will prefer higher educational standards. Assume that the utility function and the educational production function take the forms:

$$(1') \quad U = (L_i)^{0.5}(w_i)^{0.5} \quad \text{subject to } L_i \in [0, 1] \quad \text{and}$$

$$(2') \quad VMP_i = \pi_i = [2 - L_i + ka], \text{ s.t. } k > 0.$$

Note that ability and effort are perfect substitutes in the production of educational achievement. Assume further that ability a is distributed uniformly on $[0, 1]$.[8] Under these assumptions, it is easy to calculate wages for both groups of students and a^* as a function of the standard π_s. For all nonnegative values of a^*, it can be shown that $da^*/d\pi_s > 0$.

The planner's objective function is an iso-elastic function of each worker's income:

$$(11) \quad V = F(a^*)\frac{(w^2)^z}{z} + (1 - F(a^*))\frac{(w^1)^z}{z}.$$

The planner's elasticity of substitution between the income of any two workers is given by $1/1 - z$. Thus the income-maximizing planner has $z = 1$. Egalitarian planners have $z < 1$. The lower is z, the more egalitarian is the planner, and as $z \to -\infty$ we approach the case of the Rawlsian planner for whom social welfare depends only on the income of the person with the lowest income in the population.

The planner's problem was solved numerically for various values of k, where k captures

[7] Note that in the proof of Result 3, the sign of $dV^E/d\pi_s$ when evaluated at the income-maximizing planner's optimal standard is ambiguous precisely because the wage of the less-able group, w^2, varies with the standard. This will not occur if all members of the bottom group are equally productive.

[8] Note that the educational production function is chosen so that productivity of the lowest-ability worker will be positive even if he exerts no effort while in school.

the importance of ability relative to effort in determining educational achievement. The results appear in Figure 2. For each value of k, the educational standard is normalized to 1 at $z = 1$ to facilitate graphing.[9]

For $k = 1$ and $k = 10$, the more egalitarian is the planner (the lower is z), the *higher* he sets the educational standard. The effect is strongest for the case $k = 10$, where the variation in educational achievement due to innate ability is relatively high. In this case, a moderately egalitarian planner (with z of approximately 0.745) reaches a corner solution, setting an educational standard so high that $a^* = 1$. In this way he achieves complete pooling, thereby raising the earnings of the less-able students. The egalitarian planner favors the relatively large increases in w^2 that result from more-able workers dropping into the pool of students who do not meet the standard. For the case $k = 1$, where ability does not contribute as strongly to achievement, it requires a slightly more egalitarian planner to set the standard so high that nobody passes it.

Note that a planner who is so egalitarian as to prefer the pooling outcome will be indifferent as to whether we achieve this outcome by setting the standard so high that nobody can attain it, or so low that everybody can attain or exceed it with zero effort. In other words, there are two solutions to the planner's problem whenever the optimal outcome is for all students to be placed in the same group: either everybody is allowed to pass or everybody is made to fail. Although neither of these extremes is particularly realistic, there is a case to be made that the former situation is closer to the status quo in American public schools. [See for instance Lorrie A. Shepard and Mary Lee Smith (1990), who estimate that in their sample of 13 states and the District of Columbia, the average proportion of students who are held back a grade in a given year is about 6 percent.]

The most interesting case is $k = 0.5$, where there is a range in which it is true that the more egalitarian the planner, the lower the planner should set the standard. For this level of k, the differences in productivity by ability are so small that the mildly egalitarian planner sets educational standards lower than does the income-maximizing planner, to increase the proportion of the students who meet the standard. The losses in w^2 that result from the drop in average ability in group 2 are minor even from the point of view of this egalitarian planner. But as in the other two cases, there comes a point beyond which increasing the degree of egalitarianism requires the planner to set higher standards in order to guarantee higher wages for the least able in the population.

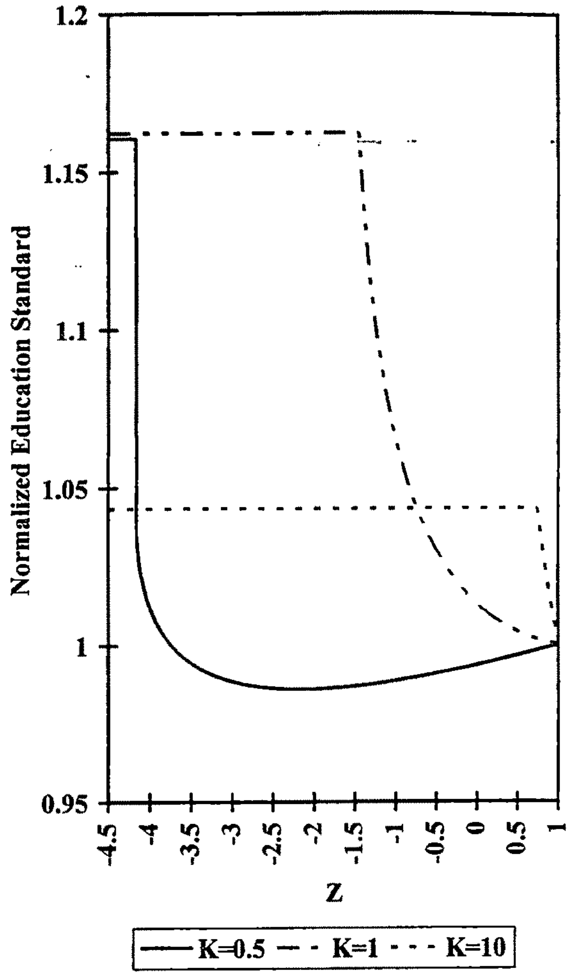


FIGURE 2. OPTIMAL EDUCATIONAL STANDARD VS. PLANNER'S PREFERENCE PARAMETER z

Notes: Optimal standard is normalized to 1 for $z = 1$. K = ability coefficient in educational production function.

This illustrates the result that regardless of the relative importance of effort and ability in producing academic achievement, as the

[9] The figure does not convey information about the relative levels of the optimal standards for different values of k because of the normalization used. As one might expect, the optimal standard rises with k. For $z = 1$, the optimal standard rises sevenfold between $k = 0.5$ and $k = 10$.

planner's preferences tend toward those of the Rawlsian planner ($z \to -\infty$), the planner will begin to raise, not lower, standards.

C. *Extension of the Social Welfare Function*

Result 3 assumes that planners are concerned with the earnings of students rather than with their utility. The next result indicates that highly similar intuition obtains for planners who maximize a function of individual utilities. In addition, a simple sufficient condition emerges under which egalitarian planners will set *higher* standards than would a planner who weights all workers equally.

RESULT 4: *Consider an egalitarian planner who maximizes a concave function of students' utility functions,* $\int_{\underline{a}}^{\bar{a}} h(U(L, w; a))f(a)\,da$. *This egalitarian planner may set a higher or lower educational standard than the "utility-maximizing" planner who maximizes* $\int_{\underline{a}}^{\bar{a}} U(L, w; a)f(a)\,da$. *Assume an interior solution for the utility-maximizing planner, so that the first-order condition is met. Then a sufficient but not necessary condition for the egalitarian planner to set a higher standard is as follows: at the optimal standard for the utility-maximizing planner, for all students of ability* $a > a^*$, *where* a^* *is the ability of the marginal worker at the optimal standard for the utility-maximizing planner,* $(dU(L^*(a, \pi_s), w^1)/d\pi_s) < 0$. *Here* $L^*(a, \pi_s)$ *is implicitly defined by* $\pi_s = g(\bar{L} - L^*, a)$, *i.e.,* L^* *is the maximum amount of leisure consistent with a person of ability a reaching the standard* π_s.

PROOF:
See the Appendix.

The intuition for the result that an egalitarian planner may set higher standards is highly similar to that for Result 3. The least-able workers see a rise in utility when the standard is raised, since their wages rise without a concomitant increase in effort. Wages rise for the bottom group because of the assumption of heterogeneous productivity among workers who exert no effort. In contrast, utility must fall with an increase in the standard for those with ability just above a^*, since in the relevant region for these workers, the indifference curves will cut the production possibilities frontier from above, as in Figure 1. (Note that for the indifferent worker with [illegible] a^*, the optimal standard would be at the [illegible] of tangency between the production possibilities frontier and the indifference curve, which [illegible] be at a lower standard than the current standard. A leftward move along the production possibilities frontier must move these workers' [illegible] lower indifference curve. Thus it is conceivable that *all* workers above a^* (or nearly all) [illegible] see their utility fall as the standard is raised, especially if a^* is near $\bar{a}$. The egalitarian planner would then set a higher standard than the utility-maximizing planner since the former discounts the utility losses to the most able while weighting heavily the utility gains that accrue to the [illegible] able as the standard is increased.[10,11]

III. Concluding Remarks

The central finding of the literature on educational standards is that higher standards [illegible] crease the achievement of only the most [illegible] students (or those with the lowest rate of [illegible] count). Thus, the literature argues, a policy maker who is concerned about inequality may want to set relatively low educational standards.

The central finding of this paper is that if workers differ in ability, and if firms cannot fully perceive individual workers' productivity, then an egalitarian policy maker might set an educational standard which is *higher* than that set by an income-maximizing planner.

This finding derives from the observation that if workers vary in ability, then an incremental increase in standards will increase the wages of

[10] For a worker with much higher ability than a^*, it becomes possible that the tangency might lie above the current standard. Such a worker will see an increase in utility if the standard is raised. On a related note, the proof of Result 4 in the Appendix shows that the sufficient but not necessary condition for the egalitarian planner to set a higher standard requires that there must be no workers above ability a^{**}. That is, all students must exert some effort to achieve the standard. Otherwise these workers would see an increase in earnings and no increase in effort, and their utility would rise with the standard.

[11] In Costrell's model, the egalitarian planner will set lower standards since w^2 is not a function of who [illegible] out." But Section II, subsection B, of his paper develops an interesting argument which suggests that an egalitarian planner might set a higher standard if he puts less weight on those students who are least willing to work hard because of "their greater capacity to enjoy leisure."

not only the most-able workers but also those of the least-able workers. Those workers who do not meet the educational standard benefit from an increase in the standard: as the pool of workers who do not graduate grows, firms correctly perceive the average quality of this group to be rising, and adjust wages upward.

In real life, this idea may have some relevance in that the stigma attached to dropping out of school is probably smaller in societies where very few people graduate from school. The stigma is smaller because firms realize that even very-able workers may fail to graduate if standards are set too high relative to the resources invested in education by schools and families.

A policy of higher educational standards will not increase the earnings of all workers, given that some students on the margin will reduce both their effort and their earnings. But the majority of students, who lie both above *and* below the critical ability level, will see increases in their earnings. Such considerations suggest that in the case of educational standards, the trade-off between equity and efficiency may not be as severe as commonly thought.

Of course, in real life a policy maker with egalitarian views may not realize the benefits that accrue to weaker students from the tightening of standards. The policy maker might mistakenly focus on the reduction in utility to the marginal student if the standard is raised, without paying sufficient attention to the benefits gained by the less-able students. In such a case, the policy maker may set standards which are too low.

APPENDIX

PROOF OF RESULT 3:

The result follows by evaluating the egalitarian planner's first-order condition at the value of π_s which solves the income-maximizing planner's problem, $\tilde{\pi}_s$. First, solve the income-maximizing planner's problem:

$$\max V^I = F(a^*)(w^2) + (1 - F(a^*))(w^1).$$

The first-order condition implicitly defines $\tilde{\pi}_s$:

$$f(a^*)(w^2(a^*) - w^1(a^*, \pi_s))\frac{da^*}{d\pi_s} + F(a^*)\frac{dw^2(a^*)}{da^*}\frac{da^*}{d\pi_s} + (1 - F(a^*))\frac{dw^1(a^*, \pi_s)}{d\pi_s} = 0,$$

where to simplify notation a^* is implicitly evaluated at $\tilde{\pi}_s$. Next, evaluate the left-hand side of the first-order condition for the egalitarian planner at $\tilde{\pi}_s$ and the corresponding value of a^*:

$$f(a^*)(h(w^2(a^*)) - h(w^1(a^*, \pi_s)))\frac{da^*}{d\pi_s} + F(a^*)h'(w^2(a^*))\frac{dw^2(a^*)}{da^*}\frac{da^*}{d\pi_s} + (1 - F(a^*))h'(w^1(a^*, \pi_s)) \times \frac{dw^1(a^*, \pi_s)}{d\pi_s}.$$

Substituting for $(1 - F(a^*))$ from the first-order condition for the income-maximizing planner's problem, and rearranging,

$$= \left\{\frac{da^*}{d\tilde{\pi}_s}\right\}\left\{ f(a^*)\underbrace{[h'(w^1(a^*, \pi_s))(w^1(a^*, \pi_s) - w^2(a^*)) - (h(w^1(a^*, \pi_s)) - h(w^2(a^*)))]}_{-}\right\}$$

$$\left\{\frac{da^*}{d\pi_s}\right\} F(a^*)\frac{dw^2(a^*)}{da^*}\underbrace{[h'(w^2(a^*)) - h'(w^1(a^*, \pi_s))]}_{+} \gtreqless 0.$$

The signs of the indicated parts of the equation are based on the facts that h is a concave function and that $w^2 < w^1$. Thus the optimal educational standard for the egalitarian planner may be above or below $\tilde{\pi}_s$.

PROOF OF RESULT 4:

The first-order condition for the planner who maximizes the sum of utilities is:

$$\frac{dV^U}{d\pi_s} = F(a^*)U_w(\bar{L}, w^2)\frac{dw^2}{da^*}\frac{da^*}{d\pi_s} + \int_{a^*}^{a^{**}} \frac{dU(L^*(a, \tilde{\pi}_s), w^1)}{d\pi_s} f(a)\,da + \frac{dU(\bar{L}, w^1)}{dw^1}\frac{dw^1}{d\pi_s}[1 - F(a^{**})] = 0.$$

In the above equation, the optimal standard for this planner is denoted by $\tilde{\pi}_s$. For the egalitarian planner the derivative of the social welfare function is:

$$\frac{dV^E}{d\pi_s} = F(a^*)h'(U(\bar{L}, w^2))U_w(\bar{L}, w^2) \times \frac{dw^2}{da^*}\frac{da^*}{d\pi_s} + \int_{a^*}^{a^{**}} h'(U(L^*(a, \pi_s), w^1)) \times \frac{dU(L^*(a, \pi_s), w^1)}{d\pi_s} f(a)\,da + h'(U(\bar{L}, w^1))\frac{dU(\bar{L}, w^1)}{dw^1}\frac{dw^1}{d\pi_s} \times [1 - F(a^{**})].$$

Evaluating the latter at the values of a^* and π_s which maximize V^U, it is found that the egalitarian planner may choose a standard which is higher or lower than the optimal solution for the nonegalitarian planner ($\tilde{\pi}_s$):

$$\left.\frac{dV^E}{d\pi_s}\right|_{\tilde{\pi}_s} = \int_{a^*}^{a^{**}} [h'(U(L^*(a, \pi_s), w^1)) - h'(U(\bar{L}, w^2))] \times \frac{dU(L^*(a, \pi_s), w^1)}{d\pi_s} f(a)\,da + [h'(U(\bar{L}, w^1)) - h'(U(\bar{L}, w^2))] \times \frac{dU(\bar{L}, w^1)}{dw^1}\frac{dw^1}{d\pi_s}[1 - F(a^{**})] \gtreqless 0.$$

The last term is negative, by concavity. The ambiguity arises because for workers with ability $a \gg a^*$ utility may increase, while for workers of ability just above a^*, utility declines with a rise in π_s. Thus, the first integral could be positive or negative due to ambiguity of the $dU(L^*, w^1)/d\pi_s$ term.

A sufficient but not necessary condition for the egalitarian planner to set a higher standard is that for all students of ability $a > a^*$, $dU(L^*, w^1)/d\pi_s < 0$. Note that this condition requires that there must be no workers above ability a^{**}, since such workers would always gain in utility from a marginal increase in π_s, as their leisure would remain unchanged while their wage would rise. Thus under the sufficient condition, the last term in the above equation vanishes. Under the sufficient but not necessary condition, then, $dV^E/d\pi_s|_{\tilde{\pi}_s} > 0$ by the above argument, and by the concavity of h and the fact that $U(L^*, w^1) > U(\bar{L}, w^2)$ for all $a > a^*$, which together ensure that the last term in the above equation is positive.

REFERENCES

Becker, William, and Rosen, Sherwin. "The Learning Effect of Assessment and Evaluation in High School." Discussion Paper No. 90-7, Economics Research Center, NORC, 1990.

Betts, Julian R. "Is There a Link between School Inputs and Earnings? Fresh Scrutiny of an Old Literature," in Gary Burtless, ed., *Does money matter? The effect of school*

sources on student achievement and adult success. Washington, DC: Brookings Institution, 1996, pp. 141–91.

Card, David and Krueger, Alan B. "Labor Market Effects of School Quality: Theory and Evidence," in Gary Burtless, ed., *Does money matter? The effect of school resources on student achievement and adult success*. Washington, DC: Brookings Institution, 1996, pp. 97–140.

Costrell, Robert M. "A Simple Model of Educational Standards." Unpublished manuscript, University of Massachusetts, 1993.

———. "A Simple Model of Educational Standards." *American Economic Review*, September 1994, *84*(4), pp. 956–71.

Guasch, J. Luis and Weiss, Andrew. "Wages as Sorting Mechanisms in Competitive Markets with Asymmetric Information: A Theory of Testing." *Review of Economic Studies*, July 1980, *47*(4), pp. 653–64.

———. "An Equilibrium Analysis of Wage-Productivity Gaps." *Review of Economic Studies*, October 1982, *44*(4), pp. 485–97.

Hanushek, Eric A. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature*, September 1986, *24*(3), pp. 1141–77.

———. "The Impact of Differential Expenditures on School Performance." *Educational Researcher*, May 1989, pp. 45–51, 62.

———. "When School Finance 'Reform' May Not be Good Policy." *Harvard Journal on Legislation*, Summer 1991, *28*(2), pp. 423–56.

———. "School Resources and Student Performance," in Gary Burtless, ed., *Does money matter? The effect of school resources on student achievement and adult success*. Washington, DC: Brookings Institution, 1996, pp. 43–73.

Kang, Suk. "A Formal Model of School Reward Systems," in John H. Bishop, ed., *Incentives, learning, and employability*. Columbus, OH: National Center for Research in Vocational Education, Ohio State University, 1985, pp. 27–38.

Shepard, Lorrie A. and Smith, Mary Lee. "Synthesis of Research on Grade Retention." *Educational Leadership*, May 1990, *47*(8), pp. 84–88.

Weiss, Andrew. "A Sorting-cum-Learning Model of Education." *Journal of Political Economy*, June 1983, *91*(3), pp. 420–42.

Competition over More Than One Prize

By DEREK J. CLARK AND CHRISTIAN RIIS *

In many situations, competition may be used as a means for providing incentives. Suppose that a firm is restricted to pay the same wage to all workers who have the same job (as is broadly the case for lecturers at Norwegian universities, for example). This situation may arise due to a union-imposed restriction, or it may be voluntary.[1] In order to provide an incentive to make the workers exert extra effort without breaking the uniform wage restriction, the firm may instigate a contest among workers with in-kind prizes such as paid leave, a family holiday, or promotion.[2] Those who exert the most effort may, on the basis of past experience for example, expect to win the prizes, but losers cannot have their effort refunded; that "bids" are sunk makes the all-pay auction a theoretical construct which can be used to model this situation.[3]

One drawback with the all-pay auction framework which limits its empirical application, however, is the concentration on the case in which a *single prize* is available (see Michael R. Baye et al., 1989, 1990; Arye L. Hillman and John G. Riley, 1989). In this paper, we extend the complete information version of the all-pay auction to allow for multiple (homogeneous) prizes. Our extension can be used at the theoretical level to examine whether established properties of the single-prize all-pay auction carry over to the more general case.

One example is the "exclusion principle" of Baye et al. (1993) in which the administrator of a single-prize all-pay auction might wish to exclude contestants with a high valuation of the prize in order to increase contest revenue. We demonstrate that this result does not necessarily hold in the several-prize case and provide intuition behind this result. Not surprisingly, one result from the single-prize case which does carry over to our more general setting is that symmetric contestants will (expectationally) completely dissipate the value of the prizes in their bidding (see Hillman and Dov Samet [1987] for the single-prize case); we show, however, that the equilibrium in this case is not necessarily symmetric.

As mentioned, one application of our multi-prize all-pay auction is to worker incentives. We examine the type of distribution mechanism, simultaneous or sequential, which a firm might use in order to stimulate extra effort in the face of a uniform wage restriction. As an example, consider the internal labor market in a firm which is looking to promote *n* workers to the next level of its hierarchy. These promotions form the prizes in an all-pay auction where those who exert the greatest effort win. In distributing these promotions, the firm has several options: promote *n* workers simultaneously, or divide up these promotions by instigating a sequential contest. If the aim is to maximize the expected extra effort exerted by workers, we show that the perceived composition of the workforce determines which of these options a firm should choose by analyzing two mechanisms. First we present a multi-prize all-pay auction in which all *n* prizes are awarded simultaneously (to the *n* workers with the highest effort); then this model is extended to cover the case in which the prizes are available sequentially (and possibly in blocks, i.e., several prizes available in each round). We characterize the unique

* Clark: Department of Economics, NFH, University of Tromsø, N-9037 Tromsø, Norway; Riis: Foundation for Research in Economics and Business Administration (SNF), Gaustadalléen 21, 0371 Oslo, Norway. Financial support from The Research Council of Norway is gratefully acknowledged. We would like to thank seminar participants at the University of Bergen, Karl Ove Moene, Michael Wallerstein, and two anonymous referees for helpful comments. Any remaining errors are our responsibility.

[1] George P. Baker et al. (1988) note the prevalence of horizontal equity in organizations.

[2] Baker et al. (1988) indicate that promotion-based incentives are common in corporate America.

[3] For a comparison of the all-pay and more traditional auction forms, see Erwin Amann and Wolfgang Leininger (1995, 1996).

equilibrium in each mechanism and calculate the expected effort of each worker as well as the probability distribution of the beneficiaries. Using an example taken from the model, we demonstrate that a firm with a dominant worker (who has a much lower cost of effort than colleagues) would maximize expected total worker effort in the contest by distributing prizes simultaneously, whereas if no worker were dominant, sequential distribution is to be preferred.

The idea of using a contest to inspire extra work effort is related to the literature on the use of a rank-order tournament as an optimal labor contract in the presence of moral hazard—see, inter alia, Edward P. Lazear and Sherwin Rosen (1981) and the extensions of their work by Jerry R. Green and Nancy L. Stokey (1983) and Barry J. Nalebuff and Joseph E. Stiglitz (1983). A common assumption in these papers is that workers have identical preferences and that worker performance is a random variable which is related to effort, which itself is unobservable; the random components of performance are drawn from a common distribution so that workers are *ex ante* identical and the focus is on a symmetric Nash equilibrium in pure strategies. This assumption is reasonable in their work since the tournament is set in a one-period (worker lifetime) framework. In contrast, our approach emphasizes differences in the participants in an environment where the rank order of worker performance can be assessed. It is reasonable to assume that contestants in an internal labor market possess some information on the abilities or type of their colleagues.[4] Those who are unsuccessful in one promotion contest may remain in the firm and compete again at a later date, giving workers a chance to assess the "types" of their colleagues. In contrast to the tournament literature in which all workers compete (using a symmetric strategy), we show that known differences between workers lead to an equilibrium in which only a subset of the workers are active in the contest.

Our results also have implications and applications in the field of rent seeking. Ever since this term was coined by Anne O. Krueger (1974), economists have been interested in the amount of resource waste which arises as a consequence of the competition for rents; Krueger's example was the competition for import licenses. Somewhat surprisingly, analysis and implications have been based almost exclusively on formal models involving a single rent.[5] Our models extend this simple case. With n identical, indivisible prizes, we demonstrate that the amount of rent seeking which arises is sensitive to whether the prizes are awarded simultaneously or sequentially. A further application of our models is to rationing by waiting in line; our results extend those of Leif Johansen (1987) who uses an all-pay auction to model a queuing situation.

The paper is organized as follows. Section I presents and analyzes the all-pay auction in which the highest n bids all win a prize, while Section II examines an extended model in which blocks of prizes may be distributed sequentially. The results of the models are applied in Section III to the case of worker incentives in the face of a nondiscriminatory wage policy. Section IV discusses further applications and implications of our results in the areas of rent seeking and goods rationing by queuing.

I. Simultaneous Distribution

Imagine an all-pay auction in which there are $n \geq 1$ identical prizes to be won. There are N players who we shall rank according to the value, v_i ($i = 1, 2, \ldots, N$), which they place upon winning a prize. The players can only

[4] The equilibrium of our complete information game can be derived as the limit of a game in which the valuations of the players are statistically independent and are private information; as the variances of the distributions from which the valuations are drawn approach zero, the equilibria of the game of incomplete information approach our mixed strategy equilibrium. This is an example of the "purification theorem" of John Harsanyi

[5] In an imperfectly discriminating rent-seeking contest, the decision rule of the contest administrator is taken to be stochastic so that players increase their probability of winning (or their share if the prize is divisible) by making rent-seeking outlays. S. Keith Berry (1993) and Clark and Riis (1996a, b) have considered this type of contest with several homogeneous, indivisible prizes and identical players.

win one prize and, as the prizes are identical, ν_i is independent of which prize is won. To simplify the model assume that $\nu_1 > \nu_2 > \cdots > \nu_N$.[6] The valuations are common knowledge. The model consists of two stages: in stage one, the N players simultaneously choose an outlay (sunk bid) $x_i \geq 0$; in stage two, the n prizes are given to the n players with the highest outlays. If two or more players bid the same amount, we assume that the probability of winning is equal for these players. We distinguish between active and passive participation by assuming that an active participant has a positive probability of bidding $x > 0$, while a passive participant will always bid $x = 0$.[7]

Let $G_i(x_i)$ represent the probability that player i wins a prize if he contributes x_i and all of the other players follow their respective equilibrium strategies. Assuming risk neutrality, the expected payoff of player i is thus

$$G_i(x_i)\nu_i - x_i. \tag{1}$$

At this stage we should note an alternative interpretation of the differences between the players as expressed by the ν_i parameters. Since (1) can be reformulated as $\nu_i[G_i(x_i) - (x_i/\nu_i)]$, $1/\nu_i$ can be seen as the marginal cost to player i of making a bid; thus we can either think of player 1 as having the highest valuation of a prize or the lowest marginal cost of making a bid.

The equilibrium of this game is necessarily in mixed strategies.[8] Let $F_i(x)$ represent the cumulative density function of player i's equilibrium mixed strategy; denote the upper support of $F_i(x)$ by ν_i^u and the lower support by ν_i^l. Proposition 1 characterizes the unique mixed strategy equilibrium for this game.

PROPOSITION 1: *There exists a uni[que] mixed strategy equilibrium of the game [in] which the $n + 1$ highest ranked players bid, $i = 1, 2, \ldots, n + 1$, from probability dis[tri]bution functions $F_i(x)$ over $[\nu_i^l, \nu_{n+1}]$, [with] common upper support $\nu_i^u = \nu^u = \nu_{n+1}$ an[d] lower supports given by*

$$\nu_{n+1}^l = 0 \tag{2}$$

$$\nu_i^l = \left[1 - \prod_{j=i}^{n}\left(\frac{\nu_j}{\nu_i}\right)\right]\nu_{n+1} \qquad i = 1, 2, \ldots, n$$

and where

$$F_i(x) = 1 - \left(\frac{\nu_i}{\prod_{j=k}^{n}\nu_j^{1/(n+1-k)}}\right) \times \left(1 - \frac{x}{\nu_{n+1}}\right)^{1/(n+1-k)} \tag{3}$$

$$i = 1, 2, \ldots, n$$

where

$$k = 1 \qquad \text{if } \nu_1^l \leq x \leq \nu_{n+1}$$

$$k = s \qquad \text{if } \nu_s^l \leq x < \nu_{s-1}^l$$

$$s = 2, 3, \ldots, n.$$

Player $n + 1$ bids $x_{n+1} > 0$ with proba[bility] ν_{n+1}/ν_n. The conditional distribution function of this player is

$$F_{n+1}(x \mid x > 0) = 1 - \left(\frac{\nu_n}{\prod_{j=k}^{n}\nu_j^{1/(n+1-k)}}\right) \times \left(1 - \frac{x}{\nu_{n+1}}\right)^{1/(n+1-k)} \tag{4}$$

The expected net surplus of actively p[artici]pating player $i = 1, 2, \ldots, n + 1$ is ν_i

[6] If two or more players have an equal valuation, this leads to multiple equilibria. For the case of a single prize, see Hillman and Samet (1987), Baye et al. (1990), and Tore Ellingsen (1991).

[7] Notice that a zero bid is considered as participation; thus simply being in the set of contestants is to be interpreted as being a (passive) participant in the all-pay auction. This is a technical assumption only; in equilibrium, the probability of winning with a zero bid is zero.

[8] This is proved in the Appendix.

PROOF:

See the Appendix.

In equilibrium, players 1, 2, ... , $n + 1$ participate actively; of these, it is only player $n + 1$ who randomizes between submitting a positive and a zero bid.[9] According to Proposition 1, this equilibrium is characterized by n portions of the interval $[0, \nu_{n+1}]$. We see from (2) that $\nu_i^l > \nu_{i+1}^l$. Player 1 chooses a bid in only the highest portion $[\nu_1^l, \nu_{n+1}]$, player 2 chooses from the two highest adjoining portions, $[\nu_2^l, \nu_1^l]$ and $[\nu_1^l, \nu_{n+1}]$. Player n chooses from the entire interval $[0, \nu_{n+1}]$ (i.e., all n portions) in equilibrium, as does player $n + 1$ conditional on submitting a strictly positive bid. We see from (3) and (4) that the probability distribution used by the players depends upon the particular portion of the interval in which the distribution is used.

We next derive the probability for each player of winning a prize in the unique equilibrium in Proposition 1. From (3) it is straightforward to show that the conditional distribution functions (conditional upon $x \geq \nu_s^l$) *are symmetric, viz.*

$$F_i(x \mid x \geq \nu_s^l) = F_s(x) \qquad \forall i > s,$$

the right-hand side of which is independent of i. We can use this symmetry to calculate the probability that a particular player wins a prize. Define q_i^s as the probability that player i does not win a prize conditional upon $x_j \geq \nu_s^l$ $\forall j = 1, 2, \ldots, n + 1$.

Consider first player 1. The probability that player one loses, given that all players $j = 1, 2, \ldots, n + 1$ choose $x_j \geq \nu_1^l$, is equal to $1/(n + 1)$, since the conditional probability functions are identical for each player. Hence q_1^1 is equal to $1/(n + 1)$. If we go one step further, we find $q_1^2 = (\nu_2/\nu_1)^n q_1^1$, since the probability that all of the other players choose $x_j \geq \nu_1^l$, given that they choose $x_j \geq \nu_2^l$, is equal to $(\nu_2/\nu_1)^n$.[10] Next, $q_1^3 = (\nu_2/\nu_1)(\nu_3/\nu_1)^{n-1} q_1^1$, since ν_2^l is player 2's lower support, thus player two chooses $x_2 > \nu_1^l$ with probability ν_2/ν_1, while the remaining players choose $x_j > \nu_1^l$ with probability ν_3/ν_1.

Consider next player i. The probability that player i loses given that all other players choose $x_j \geq \nu_i^l$ can be written

$$q_i^i = (q_1^i + q_2^i + \cdots + q_{i-1}^i)0$$

$$+ (1 - q_1^i - q_2^i - \cdots - q_{i-1}^i)\frac{1}{n + 2 - i}$$

$$= (1 - q_1^i - q_2^i - \cdots - q_{i-1}^i)\frac{1}{n + 2 - i}.$$

Taking into account the symmetry of the conditional distribution functions it follows that, for $i < s$,

$$q_i^s = \left(\frac{\nu_s}{\nu_i}\right)^{n+2-s} \prod_{j=i}^{s-1} \left(\frac{\nu_j}{\nu_i}\right) q_i^i.$$

Hence player i's probability of winning a prize when the n highest bidders are all successful, $p_i(n)$, can be written

$$(5) \qquad p_i(n) \equiv 1 - q_i^{n+1}$$

$$= 1 - \prod_{j=i}^{n+1} \left(\frac{\nu_j}{\nu_i}\right) q_i^i.$$

Given the probability of each player winning a prize, it is straightforward to calculate the equilibrium expected bids. From Proposition 1, we have that the net surplus of player i is $\nu_i - \nu_{n+1}$; however, the *gross surplus* is $p_i(n)\nu_i$. The difference between the gross and net surplus is the expected bid, $x_i(n)$. Thus

$$(6) \qquad x_i(n) = p_i(n)\nu_i - (\nu_i - \nu_{n+1}).$$

[9] Consider the usual case of $n = 1$. Players 1 and 2 participate actively; 1 submits a strictly positive bid, while player 2 has a probability ν_2/ν_1 of making a strictly positive bid. With $k = 1$ in (3), as the players both choose from the same single interval, the equilibrium strategy is to choose from the interval $[0, \nu_2]$ by use of the uniform distribution $F(x) = x/\nu_2$ (player 2 uses this distribution conditional upon submitting a strictly positive bid). These results are identical to those in Hillman and Riley (1989).

[10] To see this note that the probability that player j chooses $x_j \geq \nu_1^l$, given that he chooses $x_j \geq \nu_2^l$, is $1 - F_j(\nu_1^l / x_j \geq \nu_2^l) = \nu_2/\nu_1$.

The total of the expected bids is simply the sum of the $x_i(n)$ in (6) for all $n + 1$ active players.

From the single-prize all-pay auction of Hillman and Samet (1987), we should expect to approach full rent dissipation when the players approach symmetry in their valuations; this means simply that the total value of the expected bids approaches the total value of the prizes on offer. This is indeed true in our model. As the valuations approach a common ν, the players use the same mixed strategy in (2) and (3). Thus $p_i(n) \to n/(n+1)$ and $x_i(n) \to n\nu/(n+1)$, so that the sum of bids is expected to be $n\nu$, which is the total value of the prizes.

One result which does not carry over so easily from the single-prize case is the "exclusion principle" of Baye et al. (1993). When $n = 1$ we find $x_1(1) = \nu_2/2$ and $x_2(1) = \nu_2^2/2\nu_1$ so that the total of the expected bids is always decreasing in ν_1; this forms the intuition behind the exclusion principle in which contest revenue can, in some circumstances, be maximized by excluding the player(s) with the highest valuation from the all-pay auction. In the multi-prize framework this is not necessarily the case. As a demonstration, consider the case $n = 2$ for which we find win probabilities

$$(7) \qquad p_1(2) = 1 - \frac{\nu_2\nu_3}{3\nu_1^2};$$

$$p_2(2) = 1 - \frac{\nu_3}{2\nu_2}\left(1 - \frac{\nu_2^2}{3\nu_1^2}\right);$$

$$p_3(2) = \nu_3\left(\frac{\nu_2}{6\nu_1^2} + \frac{1}{2\nu_2}\right)$$

with corresponding individual expected bids:

$$(8) \qquad x_1(2) = \nu_3\left(1 - \frac{\nu_2}{3\nu_1}\right);$$

$$x_2(2) = \frac{\nu_3}{2}\left(1 + \frac{\nu_2^2}{3\nu_1^2}\right);$$

$$x_3(2) = \nu_3^2\left(\frac{\nu_2}{6\nu_1^2} + \frac{1}{2\nu_2}\right)$$

so that sum of expected bids is $z(2)$ where

$$(9) \qquad z(2) = \sum_{i=1}^{3} x_i(2) = \nu_3\left[1 - \frac{\nu_2}{3\nu_1} + \frac{1}{2}\left(1 + \frac{\nu_3}{\nu_2}\right)\left(1 + \frac{\nu_2^2}{3\nu_1^2}\right)\right].$$

Notice that $z(2)$ is not monotonic in the highest valuation ν_1; if $\nu_1 > (<)\ \nu_2 + \nu_3$ then $z(2)$ is increasing (decreasing) in ν_1. When $z(2)$ is increasing in ν_1 then excluding the highest-ranked contestant will not increase expected contest revenue as suggested by Baye et al. (1993) for the case of $n = 1$. The intuition behind this result is straightforward. As ν_1 increases, so does the probability, p_1, that player 1 will win one of the available prizes. As $p_1(2)$ increases, players 2 and 3 realize that they are competing against each other for the remaining prize; in the limit as $\nu_1 \to \infty$ we see from (5) and (6) that $p_1 \to 1$ and $x_1 \to \nu_{n+1}$. In the limiting case and with $n = 2$, player 1 has a sunk bid approaching the valuation of player 3 and receives a prize with certainty. Thus the competition for the second prize is between players 2 and 3; this is the case of Baye et al. (1993) which leads to the exclusion principle. Notice that $z(2)$ is increasing in ν_1 only when there is a large enough probability that player 1 will win one of the prizes; $z(2)$ is increasing in ν_1 when $\nu_1 > \nu_2 + \nu_3$, which implies that $p_1(2) > 1 - \nu_2\nu_3/[3(\nu_2 + \nu_3)^2]$. We have a form of the exclusion principle which does not rely upon the exogenous exclusion of player 1; rather the exclusion is endogenized. As long as there is a large enough probability that player 1 will be excluded from the competition for the second of the prizes (by winning the first), then it is optimal from a contest revenue-maximizing point of view to allow this player to compete.

Notice, however, that $z(2)$ may increase by excluding player 2 from the competition: $\partial z(2)/\partial\nu_2 < 0$ always. We find again that the total of expected bids $z(2)$ is increasing in the valuation of the lowest active participant; again this is due to the fact that ν_3 is the upper support of the players' probability distribution

functions. Generally, as $\nu_{n+1} \to 0$, the upper support approaches zero and thus the expected bids also approach zero.[11]

II. Sequential Distribution

The all-pay auction of Section I with simultaneous distribution of all n available prizes will not suit all applications. In this section we investigate a flexible prize structure in which blocks of prizes are made available sequentially, retaining the assumption that no player may win more than one prize. Thus the winners in each stage are eliminated from further participation. At each stage, new bids are made and previous bids have no effect on the current or future contests. As the prizes are distributed sequentially, we allow for discounting.

For comparability with the model of the previous section, let us assume that the total number of identical, indivisible prizes available is n, and that they are distributed over $r \geq 2$ rounds. Denote by n_t the number of prizes awarded at stage $t = 1, \ldots, r$, and by $1 \geq \delta > 0$ the discount factor. Obviously, the game discussed in Section I depicts the distribution of the final block of n_r prizes. The unique equilibrium of the last stage is given in Proposition 1 where the $n_r + 1$ remaining players (who have not previously won) who have the highest valuations actively participate. The n_r + 1st player randomizes between submitting a zero and a strictly positive bid; denote the valuation of this player by ν^r. Applying Proposition 1, the players who are active in the last stage receive a current expected payoff $\nu_i - \nu^r$. Consider the payoff function of player i who actively participates in stage $r - 1$,

$$(10) \quad G_i^{r-1}(x_i)\nu_i + (1 - G_i^{r-1}(x_i))\delta(\nu_i - \nu^r) - x_i$$
$$= G_i^{r-1}(x_i)[(1-\delta)\nu_i + \delta\nu^r] - x_i + \delta(\nu_i - \nu^r)$$
$$\equiv G_i^{r-1}(x_i)u_{i,r-1} - x_i + \delta(\nu_i - \nu^r).$$

Clearly, the two-stage game comprising stages $r - 1$ and r is equivalent to a one-stage game inserting $u_{i,r-1}$ for ν_i. Observe that the ranking of the players is the same in the u-series as in the ν-series. Hence, using Proposition 1, among the set of players who have not previously won, the $n_{r-1} + 1$ players with the highest valuations participate actively in stage $r - 1$ and obtain an expected payoff equal to $[(1-\delta)\nu_i + \delta\nu^r] - [(1-\delta)\nu^{r-1} + \delta\nu^r] + \delta(\nu_i - \nu^r) = (1-\delta)(\nu_i - \nu^{r-1}) + \delta(\nu_i - \nu^r)$, where ν^{r-1} is the valuation of the lowest-ranked active player in stage $r - 1$. Consider next stage t. By backward induction it follows that the payoff function of player i who is active in stage $t = 1, \ldots, r - 2$ is,

$$(11) \quad G_i^t(x_i)[(1-\delta)\nu_i + \delta\bar{\nu}^t] - x_i + \delta(\nu_i - \bar{\nu}^t)$$
$$\equiv G_i^t(x_i)u_{i,t} - x_i + \delta(\nu_i - \bar{\nu}^t)$$

$$\text{where } \bar{\nu}^t = (1-\delta)\sum_{\lambda=t}^{r-2}\delta^{\lambda-t}\nu^{\lambda+1} + \delta^{r-t-1}\nu^r.$$

It follows from Proposition 1 that the expected payoff of a player participating actively in stage t is $(1-\delta)(\nu_i - \bar{\nu}^t) + \delta(\nu_i - \nu^r)$. From this discussion and Proposition 1, the following is straightforward to show:

PROPOSITION 2: *There exists a unique mixed strategy equilibrium of the extended all-pay auction which is characterized as follows. At each stage $t = 1, 2, \ldots, r$ the $n_t + 1$ remaining players with the highest valuations all actively participate. The loser of each stage*

[11] Due to the fact that the valuation of the n + 1st player forms the upper support of all players' mixed strategy, it is not obvious that excluding player 2 will increase expected contest revenue. While $z(2)$ is decreasing in ν_2, excluding this player reduces the common upper support $(\nu_3 < \nu_2)$ which will decrease expected bids. This is the trade-off present in Baye et al. (1993).

$t = 1, \ldots, r-1$ goes on to the next round to compete against the n_{t+1} highest ranked of the players who have not yet actively participated. The equilibrium mixed strategies are as given in Proposition 1, replacing ν_i by $u_{i,t} = (1-\delta)\nu_i + \delta\bar{\nu}'$. Considered from stage 1, the expected net surplus of actively participating player $i = 1, \ldots, n+1$ is $\delta^{t(i)-1}[(1-\delta)(\nu_i - \bar{\nu}^{t(i)}) + \delta(\nu_i - \nu^r)]$, where $t(i)$ denotes the first stage at which player i actively participates.

Notice that $u_{i,t}$ is simply a weighed average (with δ as the weight) of a player's own valuation and $\bar{\nu}'$ which is common to all players. In order to compare this extended framework with that of Section I, it is useful if the n prizes have the same value in both models. Thus we concentrate on the limiting case in which $\delta \to 1$; in this case, $u_{i,t}$ is the same for all players in stages $t = 1, \ldots, r$. From Propositions 1 and 2 we find that at each stage $t = 1, \ldots, r-1$, there are $n_t + 1$ active participants who *all* have a strictly positive bid; the equilibrium bids in stages $1, \ldots, r-1$ are taken from the common probability distribution

$$F_{it}(x) = 1 - \left(1 - \frac{x}{\nu_{n+1}}\right)^{1/n_t} \tag{12}$$

over the interval $[0, \nu_{n+1}]$. Thus the mixed strategy used at any stage $1, \ldots, r-1$ is dependent upon the number of prizes awarded at each stage only, independent of the individual valuations. At stage r, there are $n_r + 1$ active participants, one of whom randomizes between submitting a zero and a positive bid.

The total expected level of bids in the first $r-1$ stages turns out to be independent of how the prizes are distributed across these stages as well as of the number of stages, i.e., independent of $\{n_t\}$, $t = 1, \ldots, r-1$. To see this, consider the expected bid by player i who actively participates in stage $t = 1, \ldots, r-1$:

$$x_{it} = \int_0^{\nu_{n+1}} x \, dF_{it}(x) = \frac{n_t}{1+n_t}\nu_{n+1}.$$

Since the number of active players is $n_t + 1$, the total of the expected bids at stage $t = 1, \ldots, r-1$ is $n_t\nu_{n+1}$; summing over the $r-1$ stages yields the total amount which is expected to be bid as $\nu_{n+1}(n - n_r)$ which depends solely on how many of the n prizes are awarded over the first $r-1$ stages.

The outcome of the equilibrium in the final stage is dependent upon the identity of the loser of the contest in round $r-1$; this player competes with the n_r highest-ranked players who have not participated actively before the final stage. As the equilibrium at each stage $t = 1, \ldots, r-1$ is symmetric, each active player has a $n_t/(n_t+1)$ chance of winning at that stage. Consequently, the probability that player i who begins competing in stage $t = 1, \ldots, r-1$ does not win a prize in any stage prior to r is ϕ_{it} where

$$\phi_{it} = \prod_{j=t}^{r-1} \frac{1}{n_j + 1}.$$

Thus a player's probability of losing all the contests in which he is active prior to the final round depends upon the round in which the player entered; this in turns depends upon the number of prizes distributed at each stage. Hence the total probability that player i who first participates actively in stage $t = 1, \ldots, r-1$ wins a prize $(1 - \phi_{it}(1 - p_i(n_r)))$ depends upon how the awards are distributed. As a consequence, this distribution also affects the total of the expected bids for the whole contest with n prizes. As in the model of Section I, the net surplus of player $i = 1, \ldots, n+1$ in the limiting equilibrium approaches $\nu_i - \nu_{n+1}$. This means that, as we approach symmetry in the players' valuations, the expected value of total bids is equal to the total value of the prizes.

One special case of the general model which is simple to analyze in the limiting equilibrium ($\delta \to 1$) involves one prize being available in each of r rounds, $n_t = 1\ \forall t = 1, \ldots, r$. According to Proposition 2, at stage $t = 1, \ldots, r-1$, the two remaining players with the highest valuations compete actively against each other with bids drawn from the uniform distribution over $[0, \nu_{n+1}]$. The final stage is identical to the usual one-prize case, with player $n+1$ competing against the loser of stage $r-1$. Consider player i's probability of winning. Since in all stages $1, \ldots, r$

equilibrium is characterized by two players bidding x_i from identical distribution functions, the probability of an active player winning a single stage is equal to 0.5. The probability of winning stage r is given by the one-stage game with one prize which player i wins with probability $(1 - 0.5(\nu_{n+1}/\nu_i))$. This gives the total probability, P_i, that i wins a prize in this version of the sequential all-pay auction as

$$P_1 = 1 - \left(\frac{1}{2}\right)^n \frac{\nu_{n+1}}{\nu_1} \tag{13}$$

$$P_i = 1 - \left(\frac{1}{2}\right)^{n+2-i} \frac{\nu_{n+1}}{\nu_i}, \qquad i = 2, \dots, n$$

and for player $n + 1$

$$P_{n+1} = n - \sum_{i=1}^{n} P_i. \tag{14}$$

The total expected bid by player i in the course of the game is $x_i = P_i\nu_i - (\nu_i - \nu_{n+1})$.

There is an interesting difference between the simultaneous and sequential models as the players' valuations approach a common value. In this case, the model of Section I with simultaneous distribution of all n prizes leads to an approximately symmetric equilibrium in which the total value of the prizes is completely dissipated. When one prize is distributed sequentially over r rounds, we achieve the same total amount of the expected bids but we find that the equilibrium in the limit is not symmetric since $P_1 \to 1 - (1/2)^n$, $P_i \to 1 - (1/2)^{n+2-i}$ for $i = 2, \dots, n + 1$. This is due to the fact that the rank of a player determines the potential number of times that the player competes when prizes are distributed sequentially. This means that the character of the equilibrium is different when a sequential mechanism is used, even if the majority of prizes are awarded simultaneously.

V. An Application to Worker Incentives

Suppose that a firm follows a nondiscriminatory wage policy at each level of its hierarchy, and wishes to inspire extra effort.[12] One way of doing this would be to instigate a winners-take-all contest of the type discussed in the preceding sections. A variety of prizes, such as paid leave or promotion to the next level of the hierarchy, may be used. The firm would require some ordinal standard by which to measure worker performance and make a ranking on the basis of which prizes are awarded. Suppose that the workers believe, based on past experience, that there is a positive relationship between extra effort (x) and the identity of the winners; our models can then be used to calculate the expected amount of extra effort and the likely beneficiaries under different distributional rules.

Recall that the ν_i parameters can be interpreted as expressing differences in the cost of effort between workers, so that the higher the ν_i the lower the cost of effort to worker i; these parameters could be seen as expressing the abilities of the workers. There is no a priori reason to assume that a firm owner or manager believes that the workforce is homogeneous. Generally, the manager will have a belief or perception about the composition of the workforce (the ν_i); on the basis of these beliefs, the manager can work out the expected amount of extra effort which will ensue in different contest forms. To illustrate this, consider the simplest possible case in which two identical prizes are available which can be distributed simultaneously or sequentially, one after the other. Given her beliefs about the ν_i, how is the manager to distribute these prizes to achieve the highest level of extra effort? Simultaneous distribution of the two prizes yields the effort levels and win probabilities in (7)–(9). If the prizes are awarded sequentially, one after the other we find (using capital letters to distinguish from simultaneous distribution)

[12] We assume that this extra effort is valuable to the firm without explicitly modelling how this affects output or profit. In contrast, "effort" in the applications mentioned in Section IV may be regarded as socially wasteful and thus the goal would be to minimize this activity.

$$P_1(1, 1) = 1 - \frac{\nu_3}{4\nu_1}; \tag{15}$$

$$P_2(1, 1) = 1 - \frac{\nu_3}{4\nu_2};$$

$$P_3(1, 1) = \frac{\nu_3}{4}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right).$$

for the win probabilities,

$$X_1(1, 1) = \frac{3\nu_3}{4}; \qquad X_2(1, 1) = \frac{3\nu_3}{4}; \tag{16}$$

$$X_3(1, 1) = \frac{\nu_3^2}{4}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)$$

for individual effort levels, and

$$Z(1, 1) = \sum_{i=1}^{3} X_i(1, 1) \tag{17}$$

$$= \nu_3\left[\frac{3}{2} + \frac{\nu_3}{4}\left(\frac{1}{\nu_1} + \frac{1}{\nu_2}\right)\right]$$

as the expected total amount of effort.

$Z(1, 1)$ increases in ν_3; however, $Z(1, 1)$ is decreasing in ν_1 and ν_2, via the effect which these variables have on the bid of player 3.[13] Player 3 competes only for the second prize, facing player 1 or 2, whoever is the loser of the first contest. The probability that player 3 gives a strictly positive bid is ν_3/ν' where ν' is the valuation of the loser of contest one; thus this probability is decreasing in ν_1 and ν_2. If player 3 bids $x_3 > 0$ then he wins the second contest with probability 0.5; the probability of bidding $x_3 > 0$ is ν_3/ν_1 if player 1 is the opponent (probability 0.5 of this occurring) and ν_3/ν_2 if player 2 is the opponent (probability 0.5). Multiplying these terms yields $P_3(1, 1)$.

The amount of effort exerted by players 1 and 2 is dependent only upon the upper support of the probability distribution function which they use in equilibrium. Players 1 and 2 compete over the first prize, each exerting an effort of $\nu_3/2$ and winning the prize [illegible] probability 0.5. Each has thus a 0.5 probability that they will lose and proceed to the second contest where player 3 is the opponent; [illegible] is the effort level in the second contest by the player who loses the first. Thus the expected amount of effort by players 1 and 2 is $3\nu_3$[illegible]

Now consider which of the two contests, simultaneous or sequential, yields the greater expected effort in total. Note that only the three workers with the lowest marginal cost of effort (highest ν_i) will be active in either contest form. Consider what happens when $\nu_1 \rightarrow \infty$ so that ν_1 is very large in relation to ν_2 and ν_3; this may be interpreted as worker 1 having a much lower cost of effort than the other two. Comparing (9) and (17) indicates that $z(2) > Z(1, 1)$ in this case so that simultaneous distribution yields the greater expected total effort. In the sequential contest the fact that relatively little effort is exerted can be explained as follows.[14] When ν_1 [illegible] worker 1 does not need to employ a strategy which would certainly ensure that he beats worker 2 in the first round, since he will certainly beat worker 3 in any second round contest. If worker 2 wins the first contest (which occurs with probability $^1/_2$), then worker 3 exerts no effort against the far-superior worker 1; thus there is a probability of $^1/_2$ that worker 3's effort is lost in the sequential setup. In the case of simultaneous distribution, the fact that worker 1 is so dominant (with an effort level approaching ν_3 as $\nu_1 \rightarrow \infty$) has the effect of making workers 2 and 3 exert effort as if they were the only contestants competing for a single prize. The effect of this is to raise the probability that 3 has a strictly positive effort compared to the corresponding probability in the sequential mechanism; this raises the expected effort level of player 3. Take now the case in which worker 1 is not so dominant: [illegible] $\nu_1 \rightarrow \nu_2$ for example, we find that $Z(1, 1)$ [illegible] $z(2)$ so that sequential distribution is to be preferred in order to achieve the greatest expected effort level. Which type of contest the [illegible] should instigate thus depends upon the perceived composition of its workforce.

[13] Thus the logic behind the exclusion principle is valid in this case.

[14] We are grateful to a referee for pointing [illegible] tuition here.

Both the simultaneous and sequential models yield an expected net surplus for player i of $\nu_i - \nu_{n+1}$; however, the individual probabilities of winning and individual effort levels are different in the two models. If the prizes are promotions, then the manager may wish to investigate the contest form which gives the largest probability that the workers of high ability win (high ν_i). Notice that only the $n + 1$ players with the highest ν_i will actively compete, whatever form the contest takes.[15] For the case $n = 2$, we find that $p_3(2) > P_3(1, 1)$, $p_2(2) < P_2(1, 1)$, $x_3(2) > X_3(1, 1)$ and $x_2(2) < X_2(1, 1)$. Thus worker 3 has a larger chance of winning (due to a higher relative effort level) in the simultaneous contest; this is because this contest affords player 3 a chance of beating two contestants whereas this player competes only once and against only one other contender in the sequential contest. Worker 2, on the other hand, has a greater chance of winning the sequential contest as he competes twice here and against only one other player each time. For worker 1, we find that

$$p_1(2) - P_1(1, 1) = \frac{\nu_3}{3\nu_1}\left(\frac{3}{4} - \frac{\nu_2}{\nu_1}\right)$$

so that $p_1(2) > P_1(1, 1)$ if $3/4 > \nu_2/\nu_1$. To see the intuition here, fix ν_1 and consider a reduction in the value of ν_2: this has no effect on the probability that worker 1 wins in the sequential contest, since workers 1 and 2 still employ a common strategy in the first round and 1 faces worker 3 should the former lose the first contest. In contrast, the fall in ν_2 causes worker 2 to reduce his relative effort in the simultaneous setup, and the end effect of this is to increase the probability that worker 1 is successful.

IV. Other Applications

A. *Rent Seeking*

The fact that a decision maker often has a degree of discretion in the granting of concessions, permits, or transfers, for example, often gives rise to activity and resource use which is explicitly intended to influence the decision [see Gary S. Becker (1983, 1985)]. Such "rent-seeking" activity has been the subject of a large literature (see Shmuel Nitzan [1994] for a survey). Since those who fail to influence the decision maker do not have their "bids" refunded, the all-pay auction has often been used to model rent-seeking activity (see Hillman and Samet [1987]; Hillman and Riley [1989]; Ellingsen [1991]). By explicitly modelling the competition for several rents, we have extended traditional rent-seeking theory which typically concentrates on the case of a single rent. We have shown that the characteristics of the equilibrium, the amount of rent seeking, and the probability distribution of the beneficiaries are sensitive to whether the rents are made available simultaneously or in some sequential manner.

In the rent-seeking tradition, our models most resemble those of Mark Gradstein and Nitzan (1989), who examine a situation in which identical players compete for a number of indivisible rents; each player must decide for which prize(s) to compete and how much to bid for each prize. With homogeneous prizes, their focus is on symmetric Nash equilibria. In contrast, our approach presents an all-pay auction with complete information in which competitors with *asymmetric valuations* bid for several homogeneous indivisible prizes; thus we do not restrict attention to symmetric equilibria.

Nitzan (1994) indicates that one of the major challenges facing rent-seeking theory is to endogenize the contest form; several writers have discussed the possibility of the contest administrator designing a contest in order to achieve the maximum contest revenue. In the context of the single-prize all-pay auction, Ian Gale and Mark Stegeman (1994) suggest the use of handicaps, while Baye et al. (1993) propose excluding some of the players. Elie Appelbaum and Eliakim Katz (1987) look at the optimal value of the prize in an imperfectly discriminating contest, while Robert Michaels (1988) and Amihai Glazer (1993) consider endogenization of other parameters in these models. Others have considered the possibility that an outsider (i.e., not the contest administrator) can influence the form of the contest. Society may

[15] This is in contrast to the tournament literature discussed earlier, where agents are identical *ex ante* and all workers compete in the symmetric Nash equilibrium.

be interested in limiting the amount of waste from rent seeking, for example. For an imperfectly discriminating contest, Avinash Dixit (1987) suggests that a method of achieving this is to dictate the order of play; Roger D. Congleton (1984) indicates that the decision-making authority in a rent-seeking contest should be delegated to a committee rather than to a single individual. It is important to note that rent seeking is not just a phenomenon which arises as a consequence of governmental or political decision-making. Paul R. Milgrom (1988) discusses rent seeking, or influence activities, in organizations; rent-seeking activity may well arise in firms with a fixed hierarchical structure as workers at each level compete among themselves for a limited number of promotions to the next level. In contrast to the effort contest in Section III as a means of securing promotion, the rent-seeking view suggests that workers spend time on nonproductive activity in the quest for promotion.

Our two models represent a reasonable way in which the contest form can be endogenized. By deciding how the prizes are to be distributed, sequentially or simultaneously, an administrator may be able to maximize contest revenue. Alternatively, an outsider may be able to limit the amount of rent seeking or increase the probability that certain players are the likely beneficiaries by choosing how the administrator is to award the prizes. It is important to note that we are not suggesting that a contest administrator (or outsider) will explicitly *choose* an all-pay auction as a distribution mechanism; rather, we feel that the all-pay auction is a useful modelling device when a decision process involves discretion so that the final decision is open to influence. Given that a process which resembles an all-pay auction arises, our models can be used to assess the effects of awarding all of the prizes simultaneously and awarding blocks of prizes sequentially.

B. *Rationing by Waiting*

Johansen (1987)[16] used the all-pay auction under complete information to analyze a situation in which a good could be ratio[illegible] waiting; bids are interpreted as arrival [illegible] that a high bid means arriving early in [illegible] secure the good. His analysis conce[illegible] mainly upon the mixed strategy equilib[illegible] which arose when a single good was [illegible] although he did also begin to consider the [illegible] in which symmetric agents compete for [illegible] homogeneous, indivisible units of the [illegible] However, he did not give a full characteriza[illegible] of the equilibrium except for some spec[illegible] cases, conjecturing that the general case [illegible] need to be solved using numerical analysis. [illegible] models which we present and solve expli[illegible] depict exactly this general case.

Charles A. Holt, Jr. and Roger Sh[illegible] (1982) consider a waiting-line auction m[illegible] in which valuations are private inform[illegible] and there is an entry fee (defined as the [illegible] of travelling to the queue). The formal [illegible] which they use is similar to our simult[illegible] model of Section I, in which each cons[illegible] may win no more than one unit of the ratio[illegible] good; since consumers are *ex ante* ident[illegible] the focus in their paper is on symmetric ([illegible] strategy) equilibrium. Our model, in con[illegible] is one of complete information in which [illegible] personal differences in valuations are im[illegible]tant for equilibrium behavior. Our results [illegible] those of Holt and Sherman (1982) could [illegible] seen as steps toward solving the general [illegible] of the waiting-line auction in which valu[illegible] are private information, but consumers are [illegible] *ex ante identical.*[17]

APPENDIX

PROOF OF PROPOSITION 1:

Denote by m the number of players tha[illegible] $x > 0$ with a positive probability (i.e., [illegible] participants). Observe first that m is at le[illegible] $n + 1$, otherwise the optimal outlays [illegible] (where $\varepsilon > 0$ is arbitrarily small), since [illegible]

[16] This article was actually written in 1982, shortly before Johansen's death; thus, his work on the all-pay auction actually predates many of the papers to which [illegible] already referred.

[17] Amann and Leininger (1996) consider a sing[illegible] all-pay auction in which two individuals compe[illegible] prove the existence and uniqueness of a Bayesia[illegible]rium when the valuations are private informatio[illegible] are drawn from different probability distributions.

each of the actively participating players receives a prize with certainty. Let us first show that the equilibrium distribution functions are continuous above zero. Assume the contrary so that player i chooses $x' > 0$ with a strictly positive probability and hence i's distribution function is discontinuous. Observe that $G_i(x') > 0$ otherwise i would lose with certainty and x' is not a best response. Consider any player $j \neq i$. It follows that $G_j^+(x')$ is strictly greater than $G_j(x')$ where $G_j^+(x') = \lim G_j(x)/x \to x'$ from above. Hence,

$$G_j(x_j)\nu_j - x_j$$

is discontinuous and jumps upwards at x'. Accordingly,

$$G_j^+(x')\nu_j - x' > G_j(x' - \delta)\nu_j - (x' - \delta)$$

for any δ in a sufficiently small interval $\delta \in [0, \tau]$. Hence no player $j \neq i$ bids in the interval $[x' - \tau, x']$, which implies that $G_i(x' - \tau) = G_i(x')$. Since player i may choose a bid strictly lower than x' with no reduction in the probability of winning, x' is not a best reply. Thus the equilibrium distribution function of player i must be continuous above zero.

Consider the upper support of the equilibrium distribution functions, and denote by $\nu^u \equiv \max \nu_i^u$. We will now show that $\nu_i^u = \nu^u$ for *at least* $n + 1$ players. Assume on the contrary that $\nu_i^u = \nu^u$ for strictly less than $n + 1$ players and denote by ν^{u2} the second-highest upper support. Since the equilibrium distribution functions are continuous above zero, it follows that $G(\nu^{u2}) = 1$; hence a player with upper support ν^u may reduce his expected bid without reducing the probability of winning.

Since the equilibrium strategies are mixed, it follows that $\nu_i^u > \nu_i^l$ for all actively participating players. Consider player i. Then

$$G_i(\nu^u)\nu_i - \nu^u \leq G_i(\nu_i^l)\nu_i - \nu_i^l$$

where the inequality is replaced by equality if $\nu^u = \nu_i^u$. Taking into account that $G_i(\nu^u) = 1$ for all i (since the distribution functions are continuous above zero) it follows that,

$$G_i(\nu_i^l) \geq 1 - \frac{\nu^u - \nu_i^l}{\nu_i} \geq 1 - \frac{\nu^u}{\nu_i} \geq 0,$$

where the latter weak inequality is replaced with a strong inequality if $\nu^u < \nu_i$. Assume that the number of actively participating players m is strictly greater than $n + 1$. Since $\nu_{i-1} > \nu_i$ and the valuation of the lowest-ranked active player is at least ν^u, it follows that $G_i(\nu_i^l) > 0$ for strictly more than n players. Since the equilibrium distribution functions are continuous above zero this implies that at least two active players bid $\nu_i^l = 0$ with a strictly positive probability, and not more than n players bid $x > 0$ with certainty, thus $G_i(0) > 0$. However, this implies that the winning probability is discontinuous at zero, $G_i^+(0) > G_i(0)$, which is incompatible with an equilibrium (any player is better off bidding an arbitrarily small amount, than bidding zero). Hence m is *at most* $n + 1$. As we have already shown that m is at least $n + 1$ it follows that exactly $n + 1$ players participate actively in equilibrium. Furthermore, as $G_i(\nu_i^l) \geq 1 - \nu^u/\nu_i$ and $\nu_i \geq \nu^u$ and $G_i(\nu_i^l) = 0$ for at least one player, it follows that ν^u is equal to the valuation of the lowest-ranked active player. Hence if there exists a nonactive player with a higher valuation than the lowest-ranked active player, the former can bid ν^u with certainty and make a strictly positive profit. Thus it is the $n + 1$ highest-ranked players who participate actively, $\nu^u = \nu_{n+1}$. Hence, setting $x_i = \nu^u = \nu_{n+1}$ and $G_i(\nu^u) = 1$ in payoff equation (1), we see that the net surplus of player i is $\nu_i - \nu_{n+1}$.

To show that the equilibrium distribution functions are strictly increasing, assume that, on the contrary, there exists an interval, say (ν_i', ν_i''), at which the probability density of player i's distribution function is zero. Since the number of active players is $n + 1$, it follows that $G_j(\nu_i') = G_j(\nu_i'')$ for all $j \neq i$. Accordingly, the density of any player is zero in this interval. However, this implies that

$$G_j(\nu_i')\nu_j - \nu_i' > G_j(\nu_i'')\nu_j - \nu_i'';$$

thus, ν_i'' is not an optimal reply and hence the equilibrium distribution functions are strictly increasing.

Since the equilibrium distribution functions are strictly increasing, we know that

$$G_i(x)\nu_i - x$$

is independent of x $\forall x \in [\nu_i^l, \nu_i^u]$.

Since $\nu_i^u = \nu_{n+1}\ \forall i$, hence $G_i(\nu_{n+1}) = 1\ \forall i$, it follows that the unique function for the probability that i wins is

$$\text{(A1)} \qquad G_i(x) = 1 - \frac{\nu_{n+1}}{\nu_i} + \frac{x}{\nu_i}$$

which implies that

$$G_\kappa(x) > G_{\kappa+1}(x)\ \forall\ \kappa = 1, \dots, n;$$

hence,

$$\nu_{n+1} > \nu_1^l > \nu_2^l > \cdots > \nu_n^l \geq \nu_{n+1}^l,$$

where the latter weak inequality is due to the fact that $G_i(0) = 0$, (that is, only one of the $n + 1$ active players may play an atom at zero).

Consider the interval $[\nu_1^l, \nu_{n+1}]$. Since all players bid with positive probability in this interval, player j's probability of winning, $G_j(x)$, can be written

$$\text{(A2)} \qquad G_j(x) = 1 - \prod_{i \neq j} (1 - F_i(x)).$$

From (A1) we obtain the probability that j wins (by a change of subscript), which when inserted into (A2) gives

$$\frac{\nu_{n+1} - x}{\nu_j} = \prod_{i \neq j} (1 - F_i(x)) .$$

This gives $n + 1$ equations which determine a unique solution for $F_i(x)$, $i = 1, 2, \dots, n + 1$, over the interval $[\nu_1^l, \nu_{n+1}]$. The equilibrium distribution functions over the intervals $[\nu_i^l, \nu_{i-1}^l]$, $i = 2, 3, \dots, n + 1$ are determined recursively.

REFERENCES

Amann, Erwin and Leininger, Wolfgang. "Expected Revenue of All-Pay and First-Price Sealed-Bid Auctions with Affiliated Signals." *Journal of Economics*, 1995, *61*(3), pp. 273–79.

______. "Asymmetric All-Pay Auctions with Incomplete Information: The Two-Player Case." *Games and Economic Behavior*, May 1996, *14*(1), pp. 1–18.

Appelbaum, Elie and Katz, Eliakim. "Seeking Rents by Setting Rents: The Political Economy of Rent Seeking." *Economic Journal*, September 1987, *97*(387), pp. 685–99.

Baker, George P.; Jensen, Michael C. and Murphy, Kevin J. "Compensation and Incentives: Practice vs. Theory." *Journal of Finance*, July 1988, *43*(3), pp. 593–616.

Baye, Michael R.; Kovenock, Dan and de Vries, Casper G. "The Economics of All-Pay, Winner-Take-All Contests." Working Paper No. 89-21, Texas A&M University, 1989.

______. "The All-Pay Auction with Complete Information." CentER Working Paper No. 9051, Tilburg University, 1990.

______. "Rigging the Lobbying Process: An Application of the All-Pay Auction." *American Economic Review*, March 1993, *83*(1), pp. 289–94.

Becker, Gary S. "A Theory of Competition Among Pressure Groups for Political Influence." *Quarterly Journal of Economics*, August 1983, *98*(3), pp. 371–400.

______. "Public Policies, Pressure Groups, and Dead Weight Costs." *Journal of Public Economics*, December 1985, *28*(3), pp. 329–47.

Berry, S. Keith. "Rent-Seeking with Multiple Winners." *Public Choice*, October 1993, *77*(2), pp. 437–43.

Clark, Derek J. and Riis, Christian. "On the Design of Symmetric Rent-Seeking Contests with Several Winners." Discussion Paper No. E3/96, University of Tromsø, Norway, February 1996a.

______. "A Multi-Winner Nested Rent-Seeking Contest." *Public Choice*, April 1996b, *87*(1–2), pp. 177–84.

Congleton, Roger D. "Committees and Rent-Seeking Effort." *Journal of Public Economics*, November 1984, *25*(1–2), pp. 197–209.

Dixit, Avinash. "Strategic Behavior in Contests." *American Economic Review*, December 1987, *77*(5), pp. 891–98.

Ellingsen, Tore. "Strategic Buyers and the Social Cost of Monopoly." *American Economic Review*, June 1991, *81*(3), pp. 648–57.

Gale, Ian and Stegeman, Mark. "Exclusion in All-Pay Auctions." Working Paper No. 9401, Federal Reserve Bank of Cleveland, 1994.

Glazer, Amihai. "On the Incentives to Establish and Play Political Rent-Seeking Games." *Public Choice*, February 1993, *75*(2), pp. 139–48.

Gradstein, Mark and Nitzan, Shmuel. "Advantageous Multiple Rent Seeking." *Mathematical and Computer Modelling*, 1989, *12*(4–5), pp. 511–18.

Green, Jerry R. and Stokey, Nancy L. "A Comparison of Tournaments and Contracts." *Journal of Political Economy*, June 1983, *91*(3), pp. 349–64.

Harsanyi, John. "Games With Randomly Perturbed Payoffs: A New Rationale for Mixed-Strategy Equilibrium Points." *International Journal of Game Theory*, 1973, *2*, pp. 1–23.

Hillman, Arye L. and Riley, John G. "Politically Contestable Rents and Transfers." *Economics and Politics*, Spring 1989, *1*(1), pp. 17–39.

Hillman, Arye L. and Samet, Dov. "Dissipation of Contestable Rents by Small Numbers of Contenders." *Public Choice*, 1987, *54*(1), pp. 63–82.

Holt, Charles A., Jr. and Sherman, Roger. "Waiting-Line Auctions." *Journal of Political Economy*, April 1982, *90*(2), pp. 280–94.

Johansen, Leif. "Queues (and "Rent-Seeking") as Non-Cooperative Games, Emphasizing Mixed Strategy Solutions," in Finn Førsund, ed., *Collected works of Leif Johansen*, Vol. II. Amsterdam: North-Holland, 1987.

Krueger, Anne O. "The Political Economy of the Rent-Seeking Society." *American Economic Review*, June 1974, *64*(3), pp. 291–303.

Lazear, Edward P. and Rosen, Sherwin. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy*, October 1981, *89*(5), pp. 841–64.

Michaels, Robert. "The Design of Rent-Seeking Competitions." *Public Choice*, January 1988, *56*(1), pp. 17–29.

Milgrom, Paul R. "Employment Contracts, Influence Activities, and Efficient Organization Design." *Journal of Political Economy*, February 1988, *96*(1), pp. 42–60.

Nalebuff, Barry J. and Stiglitz, Joseph E. "Prizes and Incentives: Towards a General Theory of Compensation and Competition." *Bell Journal of Economics*, Spring 1983, *14*(1), pp. 21–43.

Nitzan, Shmuel. "Modelling Rent-Seeking Contests." *European Journal of Political Economy*, May 1994, *10*(1), pp. 41–60.

Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises

By LYNNE G. ZUCKER, MICHAEL R. DARBY, AND MARILYNN B. BREWER*

The number of American firms actively using biotechnology grew rapidly from nonexistent to over 700 in less than two decades, transforming the nature of the pharmaceutical industry and significantly impacting food processing, brewing, and agriculture, as well as other industries. Here we demonstrate empirically that the commercialization of this technology is essentially intertwined with the development of the underlying science in a way which illustrates the significance in practice of the localized spillovers concept in the agglomeration literature and of the tacit knowledge concept in the information literature. Indeed we present here strong evidence that the timing and location of initial usage by both new dedicated biotechnology firms (*"entrants"*) and new biotech subunits of existing firms (*"incumbents"*) are primarily explained by the presence at a particular time and place of scientists who are actively contributing to the basic science as represented by publications reporting genetic-sequence discoveries in academic journals.

By quantifying separable effects of individual scientists, major universities, and federal research support we provide specific structure to the role of universities and their faculties in encouraging local economic development through what are conventionally described in the literature as geographically localized knowledge spillovers.[1] Such localized knowledge spillovers may play fundamental roles in both economic agglomeration and endogenous growth (Paul M. Romer, 1986, 1990; Gene M. Grossman and Elhanan Helpman, 1991). However, our evidence, like the other literature cited here, specifically indicates localized effects without demonstrating that they can be characterized as spillovers (or externalities).

Section I lays out our basic hypothesis. The data are described in Section II. Empirical results are reported and discussed in Section III. A summary and conclusions section (Section IV) and Data Appendix complete the article.

I. The Hypothesis

Innovations are generally treated in the growth literature as a nonrivalrous good—freely useable by an unlimited number of potential

* Zucker: Department of Sociology and Institute for Social Science Research, University of California, Box 951484, Los Angeles, CA 90095, and National Bureau of Economic Research; Darby: John E. Anderson Graduate School of Management, University of California, Box 951481, Los Angeles, CA 90095, UCLA Department of Economics, and National Bureau of Economic Research; Brewer: Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210. This research has been supported by grants from the National Science Foundation (SES 9012925), the University of California Systemwide Biotechnology Research and Education Program, the University of California's Pacific Rim Research Program, the UCLA Center for American Politics and Public Policy, and the UCLA Institute of Industrial Relations. We acknowledge very useful comments on earlier drafts from two anonymous referees, and from David Butz, Harold Demsetz, Robert Drazin, Martin Feldstein, Zvi Griliches, Keith Head, Adam Jaffe, Benjamin Klein, Josh Lerner, Gary Pisano, Jeff Rosensweig, L. G. Thomas, Ivo Welch, and others. We are indebted to a remarkably talented team of postdoctoral fellows Zhong Deng, Julia Liebeskind, and Yusheng Peng, and research assistants Paul J. Alapat, Jeff Armstrong, Lynda J. Kim, Amalya Oliver, Alan Paul, and Maximo Torero. Armstrong was principally responsible for conducting the analysis and cleaning the firm data set and Torero cleaned the scientist data set; comments from both substantially improved the paper. This paper is a part of the NBER's research program in Productivity. Any opinions expressed are those of the authors and not those of their employers or funders.

[1] Zvi Griliches (1992) has surveyed the importance of R&D spillovers as a major source of endogenous growth in recent "new growth theory" models and the difficulty of empirical search for their existence. Despite these difficulties, there have been a number of articles reporting evidence of geographic localization of knowledge spillovers, including Adam B. Jaffe (1989), Jaffe et al. (1993), and Edwin Mansfield (1995).

costs at a zero marginal cost (Richard R. Nelson and Romer, 1996). A complementary literature recognizes that some information requires an investment of considerable time and effort to master. The human capital developed by this investment is seen as earning a normal return on the cost of the investment, both direct costs and the cost of forgone earnings. We believe that some innovations, particularly a breakthrough "invention of a method of inventing" (Griliches, 1957), may be better characterized as creating (rivalrous) human capital—intellectual human capital—characterized by natural excludability as opposed to a set of instructions for combining inputs and outputs which can be protected only by intellectual property rights. This natural excludability arises from the complexity or tacitness of the information required to practice the innovation (see Nelson [1959], Kenneth J. Arrow [1962], Nelson and Sidney G. Winter [1982], and Nathan Rosenberg [1982]).

Based on both extensive interviews and empirical work summarized in Zucker and Darby (1996), we believe that, at least for the first 10 or 15 years, the innovations which underlie biotechnology are properly analyzed in terms of naturally excludable knowledge held by a small initial group of discoverers, their coworkers, and others who learned the knowledge from working at the bench-science level with those possessing the requisite know-how. Ultimately the knowledge spread sufficiently widely to become part of routine science which could be learned at any major research university. After the initial 1973 discovery by Stanley Cohen and Herbert Boyer of the basic technique for recombinant DNA—the foundation of commercial biotechnology as well as of a burst of scientific innovation—the financial returns available to talented recombinant-DNA scientists first rose dramatically as the commercial implications became widely appreciated and then more gradually declined as more and more scientists learned the techniques, until knowledge of the new techniques per se earned only the normal return for the time required for a graduate student to master them. Further, mere knowledge of the techniques of recombinant DNA was not enough to earn these extraordinary returns; the knowledge was far more productive when embodied in a scientist with the genius and vision to continuously innovate and define the research frontier and apply the new research techniques in the most promising areas.

We hypothesize that entry of firms into biotechnology in a given year thus will be determined by the geographic distribution of stars and perhaps others then actively practicing the new science as well as by the geographic distribution of economic activity. Stars are properly viewed as locationally (semi-)fixed since few star scientists who knew how to do recombinant DNA were willing to abandon their university appointments and laboratory teams to pursue commercial applications of biotechnology. The primary pattern in the development of the industry involved one or more scientist-entrepreneurs who remained on the faculty while establishing a business on the side—businesses which, where successful, resulted in millions or even billions of dollars for the professors who acquired early ownership stakes. Thus, we see the university as bringing about local industrial benefits by permitting its professors to pursue private commercial interests while their faculty appointments tie them to the area. In preliminary work not reported here, we tried to develop measures of local economic activity for industries, like pharmaceuticals, specifically impacted by the new technology, but these attempts never added significantly to the measures of general activity used in the empirical work below. The *local* availability of venture capital is widely believed to play a significant role in the birth of new biotech entrants (Martin Kenney, 1986; Joshua Lerner, 1994, 1995); so we also include that variable in our regressions.

II. The Data

Data has been collected in panel form for 14 years (1976–1989) and 183 regions (functional economic areas as defined by the U.S. Department of Commerce, Bureau of Economic Analysis [BEA], 1992b). Frequently, the data are aggregates of data at the zip code or county level.[2] Lagged variables

[2] The BEA's functional economic areas divide all the counties in the United States into regions including one or more cities, their suburbs, and the rural counties most closely tied to the central city.

include data for 1975 in the unlagged form. See the Data Appendix for more details.

A. *Firms*

Our data set on firms was derived from a base purchased from the North Carolina Biotechnology Center (NCBC) (1992) which was cleaned and supplemented with information in *Bioscan* (1989–1993) and its precursor (Cetus Corp., 1988). We identified 751 distinct U.S. firms for which we could determine a zip code and a date of initial use of biotechnology. Of these 751 firms, 511 were entrants, 150 incumbents, and 90 (including 18 joint ventures) could not be definitively classified. By 1990, 52 of the 751 firms had died or merged into other firms.

We then calculated the number of births in each region by year of initial use of biotechnology for all 751 firms as well as for their identified subcomponents of entrants and incumbents. We also have the stocks of surviving firms, entrants, and incumbents by region and year.

B. *Scientists*

Early in our ongoing project studying the scientific development and diffusion of biotechnology, we identified a set of 327 star scientists based on their outstanding productivity through early 1990. The primary criterion for selections was the discovery of more than 40 genetic sequences as reported in *GenBank* (1990) through April 1990.[3] However, 22 scientists were included based on writing 20 or more articles, each reporting one or more genetic-sequence discoveries.[4] In the 1990's, sequence discovery has become routin[illegible] is no longer such a useful measure of re[illegible] success. These 327 stars were only [illegible] quarters of one percent of the auth[illegible] *GenBank* (1990) but accounted for 17[illegible] cent of the published articles, almost 22 [illegible] as many articles as the average scientis[illegible]

We collected by hand the 4,061 artic[illegible] thored by stars and listed in *GenBank* [illegible] corded the institutional affiliation of th[illegible] and their coauthors on each of these [illegible] These coauthors are called "*collaborat*[illegible] they are not themselves a star. Some [illegible] the stars and collaborators who ever pub[illegible] in the United States is given on the left [illegible] Table 1, where the scientists are identifi[illegible] the organization(s) with which they [illegible] filiated on their first-such publicatio[illegible] higher citation rate for firm-affiliated sci[illegible] is explored at length in Zucker and [illegible] (1996).

Figure 1 illustrates the time pattern [illegible] growth in the numbers of stars and colla[illegible]rators who have ever published and the [illegible] number of firms using biotechnology [illegible] United States. There was a handful of [illegible] who published articles reporting gene[illegible] sequence discoveries before the 1973 [illegible] through, but even after 1973 their num[illegible] increased gradually until taking off in [illegible] The numbers of collaborators and firms [illegible] behind the growth in stars by some year[illegible]

To identify those scientists clearly wor[illegible] at the edge of the science in a given yea[illegible] term a star or collaborator as "*active*" [illegible] or she has published three or more seq[illegible] discovery articles in the three-year m[illegible] window ending with that year. As seen [illegible] right side of Table 1, this stringent [illegible] screen provides an even more elite de[illegible] of star scientists as well as identifying [illegible] very significant collaborators. We co[illegible] each year the number of active stars and [illegible] collaborators who are affiliated with [illegible]nization in each region.

The locations of active stars and fir[illegible] both concentrated and highly correlat[illegible] graphically, particularly early in the [illegible] Figure 2 illustrates this pattern for the [illegible] period by accumulating the number of [illegible] who have ever been active in each re[illegible] to 1990 and plotting them together [illegible]

[3] See Zucker et al. (1993). As will be obvious, much of the time between 1990 and the initial submission of this paper was spent in developing reasonable measures of intellectual human capital and in collecting and coding data necessary to locate the authors of the discoveries reported in the articles in question and to trace the diffusion process.

[4] Scientists advised that some sequence discoveries are more difficult than others and thus merit an article reporting only one sequence. Therefore we included scientists with 20 or more discovery articles to avoid excluding scientists who specialized in more difficult problems.

TABLE 1—DISTRIBUTION OF STAR SCIENTISTS AND COLLABORATORS WHO HAVE EVER PUBLISHED IN THE UNITED STATES

	Full data set		Ever active in U.S.[b]	
Organization type[a]	Number of scientists	Citations[c]/ scientist/years	Number of scientists	Citations[c]/ scientist/years
Stars:				
University	158	85.5	108	110.8
Institute	44	63.0	26	98.7
Firm	5	143.7	1	694.3
Dual	0	n/a	0	n/a
Total	207		135	
Collaborators:				
University	2901	10.4	369	30.6
Institute	776	13.7	88	35.8
Firm	324	29.2	43	99.1
Dual	3	7.2	0	n/a
Total	4004		500	

[a] The organization type refers to the affiliation listed on their **first** publication with a U.S. affiliation.
[b] Ever active in the U.S. means that in at least one three-year period beginning 1974 or later and ending 1989 or earlier, the scientist was listed on at least three articles appearing in our data set of 4,061 articles which reported genetic-sequence discoveries and were published in major journals and that the affiliation listed in the last of the three articles was located in the United States.
[c] Citation counts are for 1982, 1987, and 1992 for all articles in our data set (whenever published) for which the individual was listed as an author.

location of biotech-using firms as of early 1990.

C. Other Measures of Intellectual Human Capital

Active stars and collaborators may be incomplete measures of location of the scientific base because there are techniques other than recombinant DNA which have played an important role in commercial biotechnology. Some skeptical readers might also think that some simpler measures of regions' relevant academic resources would contain all the information which we have laboriously collected. We found two measures of regional scientific base which entered separately in regressions reported below, but none which were capable of eliminating the effects of the star scientists.

One measure is a count of the number of "*top-quality universities*" in a region where top quality is defined by having one or more "*biotech-relevant*" (biochemistry, cellular/molecular biology, and microbiology) departments with scholarly quality reputational ratings of 4.0 or higher in the 1982 National Research Council survey (Lyle V. Jones et al., 1982). There are 20 such universities in the United States.[5] Our second measure, "*federal*

[5] The 20 universities were: Brandeis University, California Institute of Technology, Columbia University, Cornell University, Duke University, Harvard University, Johns Hopkins University, Massachusetts Institute of Technology, Rockefeller University, Stanford University, University of California-Berkeley, University of California-Los Angeles, University of California-San Diego, University of California-San Francisco, University of

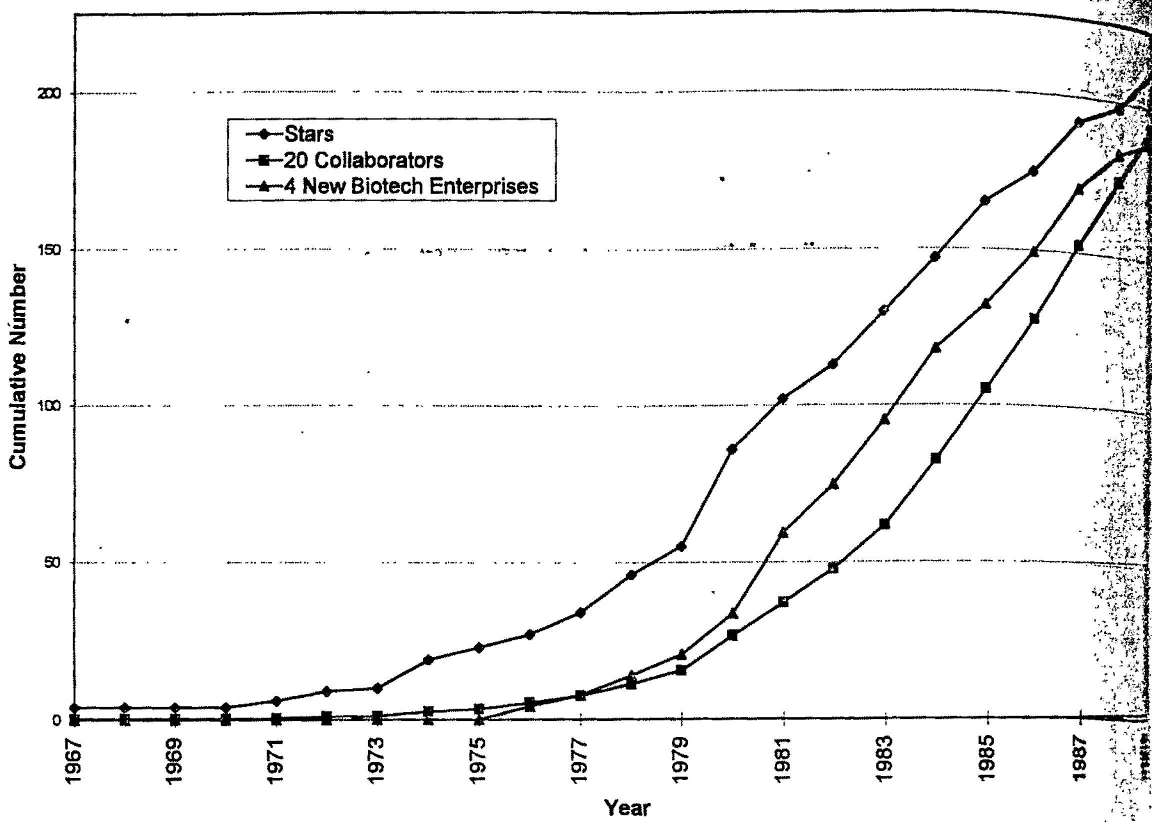


FIGURE 1. CUMULATIVE NUMBER OF U.S. STARS, COLLABORATORS, AND NEW BIOTECH ENTERPRISES, 1967–1989

support, " is the total number (in hundreds) of faculty supported by 1979–1980 federal grants to all universities in each region for biotech-relevant research.[6] These variables take on the same value for a given region in each year.

D. *Other Variables*

Using listings in Stanley E. Pratt (1982), we measure "*venture capital firms*" as the number of such firms in a region legally eligible to finance start-ups in each year up to 1981. For later years, the number of firms is fixed at the number in 1981 to avoid possible simultaneity problems once the major wave of biotech founding began.[7] (While great bookstores spring up around great universities, the former should not be counted as causing the latter.)

Since entry of biotech firms would be expected to occur where there is other economic activity, particularly involving a highly skilled labor force, we also include *total employment* in all industries (in millions of persons) and *average wages* (measured by deflated average earnings per job in thousands of 1987 dollars) for each region and year.

Finally, an increase in the (all-equity) cost of capital, as measured by the *earnings-price ratio* on the Standard & Poor's 500 index would reduce the net present value of

Chicago, University of Colorado at Denver, University of Pennsylvania, University of Washington (Seattle), University of Wisconsin-Madison, and Yale University.

[6] We also tried a measure of biotech-relevant research expenditures as reported by the universities, but this variable was too collinear with the federal support variable to enter separately and appeared to be less consistently measured across universities.

[7] Instrumental variables would provide a more ... approach to this problem if suitable instruments ... found.

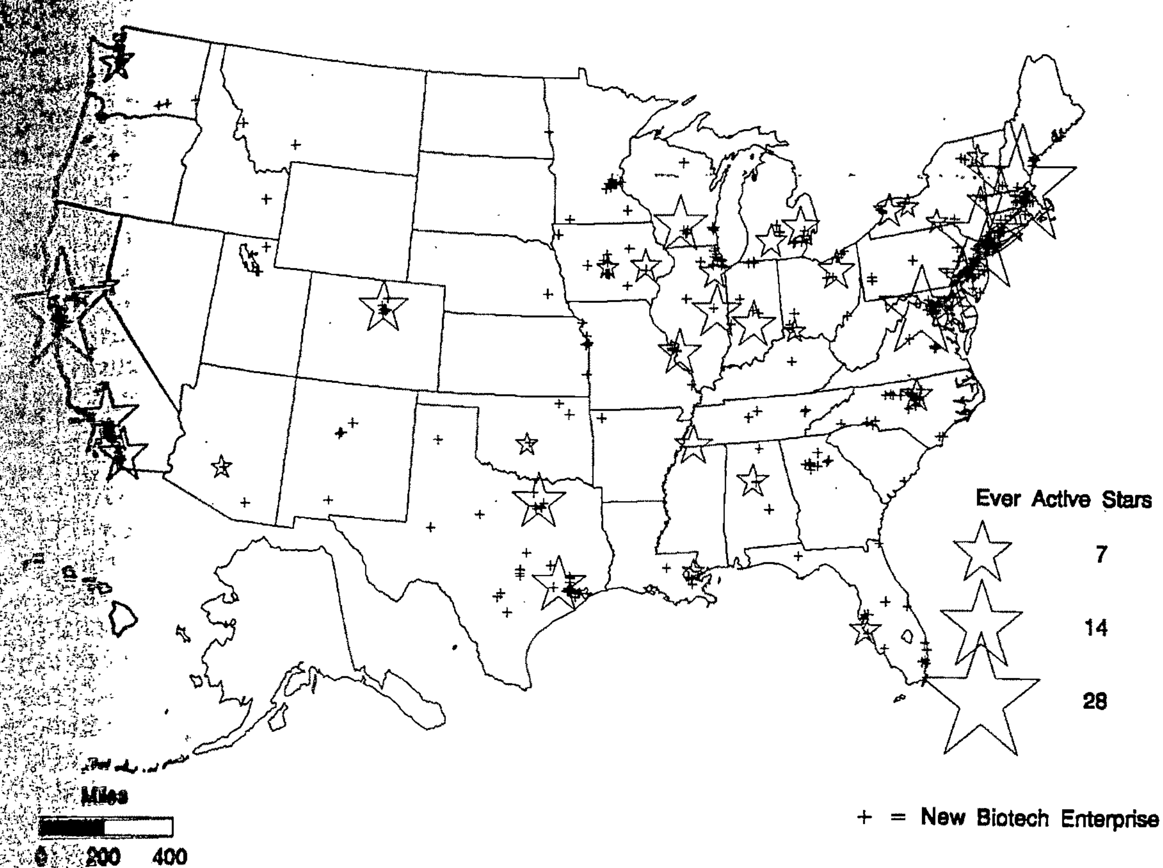


FIGURE 2. EVER ACTIVE STARS AND NEW BIOTECHNOLOGY ENTERPRISES AS OF 1990

and so should have a negative impact on birth of new firms, entrants or incumbents.

III. Empirical Results

We test our hypothesis using both the full panel data and by regressing the geographical distribution of the data in 1990 on values of the independent variables circa 1980. The former more fully exploit the available information while the latter avoid problems of possible simultaneity which might arise after 1980 when commercial biotechnology became a significant economic factor in some areas. All the regressions reported here, as well as an extensive sensitivity analysis noted below, were estimated in the poisson form appropriate for count variables with numerous zeros using LIMDEP (William H. Greene, 1992, pp. 539–49), with the Wooldridge regression-based correction for the variance-covariance matrix estimates.[8] The poisson regressions estimate the logarithm of the expected number of firm births; so the signs and significance of coefficients have the usual interpretation. Although OLS regressions are inappropriate for our count dependent variables with most observations at zero and the rest tailing off through small positive integers, we

[8] As discussed in Jerry Hausman et al. (1984), the poisson process is the most appropriate statistical model for count data such as ours. In practice, overdispersion (possibly due to unobserved heterogeneity) frequently occurs. Given the problems with resort to the negative binomial (A. Colin Cameron and Pravin K. Trivedi, 1990), Jeffrey M. Wooldridge (1991) developed a flexible and consistent method for correcting the poisson variance-covariance matrix estimates regardless of the underlying relationship between the mean and variance. We are indebted to Wooldridge and Greene for advice in implementing the procedure in LIMDEP.

reported broadly consistent results using that technique in an earlier version of the paper (Zucker et al., 1994).

In our sensitivity analysis, we ran the same poisson regressions for entrants and incumbents defined both exclusive and inclusive of the arguable case of joint ventures. The results were generally very similar to the subcomponent regressions in Table 4. In other unreported poisson regressions, we found that eliminating those regions with no firms and no stars from the sample did not result in qualitatively different results.

A. *The Long-Run Model*

Because of concerns about possible simultaneity biases once the industry became a significant economic force, we begin our empirical discussion with models which relate the number of firms in each region at the beginning of 1990 to the distribution of intellectual human capital and other variables as of about 1980. These results provide something of an acid test of our approach.

In Table 2, we present cross-section poisson regressions across the 183 regions explaining the number of firms in each at the beginning of 1990 when our data set ends.[9] Column (a) restrains the analysis to only the numbers of stars and collaborators ever active in each region at any time up through 1980, while columns (b) and (c) add first their squares and then our other intellectual human capital variables. Column (d) considers alternatively other economic variables which might explain entry, and column (e) combines the variables in (c) and (d). Column (f) adds to this model the number of biotech firms existing in 1980.

Column (a) in Table 2 indicates that the number of stars and collaborators active through 1980 is a powerful predictor of the geographic distribution of biotech enterprises in 1990, since the log-likelihood increases to −871.9 compared to −1401.7 for a constant alone. It is the star scientists that contribute positively, with collaborators having a much smaller negative coefficient in this regression and most of the other long-run models discussed below. We had expected that the coefficient on collaborators would be much smaller than that on stars, but positive. We do obtain a positive coefficient on active collaborators when the squared terms are added in column (b), but that turns negative again with addition of other variables in the remaining columns of Table 2.[10] (In the annual regressions discussed below, we generally estimate positive effects of active collaborators, but they are often statistically insignificant.)

We can offer two explanations for the generally negative sign on the number of active collaborators in the long-run regressions: (i) This coefficient reflects two partially offsetting influences; collaborators have a positive direct effect on the entry of firms but reduce the effect of stars who are devoting more of their time to training students and relatively less in starting their own firms. Training collaborators is surely a useful and rewarded activity—particularly for the academic stars—but it may take more of the stars' energy than it is worth if firm birth were the only criteria.[11] (ii) The sign and magnitude of the coefficient on collaborators may simply reflect significant multicollinearity among the intellectual human capital variables in the very early years. This is especially likely since when we examine the

[9] In an earlier version of this paper we included an alternative form of Table 2 (available from the authors upon request) in order to forestall interpretations that the results in Table 2 may reflect reverse causality. This alternative table reported regressions which explain the number of firms alive at the beginning of 1990 that were born after 1980. Nearly identical results were obtained, reflecting the fact that bulk of new biotechnology enterprises were founded after 1980.

[10] In column (b) of Table 2 (and Table 3), the negative coefficient on the squared term indicates that as the number of stars or collaborators increases, their marginal contribution diminishes eventually passing through zero. For collaborators, in columns (c)–(f) of these tables the pattern reverses so that the partial derivative of the probability of birth with respect to collaborators starts negative, and increases as their number increases, eventually becoming positive.

[11] In support of this explanation, we note that in our sensitivity analysis we tried regressions which substitute interaction terms multiplying the numbers of active stars and collaborators for the squared terms. In those regressions, we obtain significant positive coefficients on the numbers of stars and collaborators and a significant negative coefficient on their interaction.

TABLE 2—POISSON REGRESSIONS ON THE STOCK OF BIOTECH-USING FIRMS AT THE BEGINNING OF 1990 BY REGION

	(a)	(b)	(c)	(d)	(e)	(f)
Constant	0.911***	0.644***	0.468***	−2.595***	−2.718***	−2.607***
	(0.014)	(0.015)	(0.033)	(0.086)	(0.256)	(0.345)
Number stars active at any time during 1976–80	0.567***	0.587***	0.466***	—	0.877***	0.649***
	(0.029)	(0.072)	(0.090)	—	(0.076)	(0.084)
Number collaborators active at any time during 1976–80	−0.076***	0.175***	−0.183**	—	−0.333***	−0.261***
	(0.012)	(0.033)	(0.068)	—	(0.045)	(0.037)
(Number stars active at any time during 1976–80)^2	—	−0.028***	−0.019	—	−0.049***	−0.024*
	—	(0.007)	(0.014)	—	(0.012)	(0.012)
(Number collaborators active at any time during 1976–80)^2	—	−0.005***	0.002	—	0.007**	0.001
	—	(0.001)	(0.003)	—	(0.003)	(0.002)
Number top-quality universities in the region	—	—	1.388***	—	1.594***	0.442*
	—	—	(0.150)	—	(0.107)	(0.195)
Number faculty with federal support in the region	—	—	0.263	—	0.752***	0.711***
	—	—	(0.143)	—	(0.088)	(0.051)
Number venture capital firms in the region in 1990	—	—	—	0.017***	−0.045***	−0.013**
	—	—	—	(0.002)	(0.003)	(0.004)
Total employment (all industries) in the region in 1990	—	—	—	0.222***	−0.009	−0.213***
	—	—	—	(0.019)	(0.043)	(0.049)
Average wages per job in the region in 1980	—	—	—	0.166***	0.143***	0.139***
	—	—	—	(0.004)	(0.014)	(0.019)
Cumulative births of biotech firms during 1976–80	—	—	—	—	—	0.300***
	—	—	—	—	—	(0.025)
Log-likelihood	−871.9	−707.3	−543.2	−753.9	−416.0	−350.7
Restricted log-likelihood	−1401.7	−1401.7	−1401.7	−1401.7	−1401.7	−1401.7

Notes: $N = 183$. Standard errors (adjusted by Wooldridge, 1991 Procedure 2.1) are in parentheses below coefficients.
* Significantly different from 0 at the 5-percent level.
** Significantly different from 0 at the 1-percent level.
*** Significantly different from 0 at the 0.1-percent level.

full cross-section/time-series results just below we obtain (we think more reliable) zero or positive coefficients on collaborators, so the puzzle largely disappears.

The full "fundamentals" model (excepting the decade-lagged dependent variable) is presented in column (e) of Table 2, where all the coefficients are significant except that for total employment. Leaving aside the question of the negative collaborator coefficient, we note the strong positive, separate effects of stars, top-quality universities, and federal research grants at universities on birth of firms in a given geographic region. The intellectual human capital variables alone increase the log-likelihood ratio from −1401.7 to −543.2 [see column (c)], with the final three variables bringing this quantity up to −416.0. As to the last three variables, the quality of the labor force, measured by average wages per job, seems much more relevant than its size. Surprisingly, to some observers, the number of venture capital firms in a region enters, but with a significantly negative sign. We interpret

the negative sign as evidence that venture capitalists did play an active role in the formation of entrant firms, but they apparently resulted in fewer, larger firms being born in the areas in which they were more active.[12]

This sign of the coefficient of the number of venture capital firms in a region is robust in sensitivity experiments with other forms (not reported here) except for regressions which exclude the intellectual human capital variables such as in column (d). That regression looks good in terms of significance and expected sign pattern although it has a much lower explanatory power than the intellectual human capital variables alone [column (c)]. Just below, we report very similar results in a cross-section/time-series context. Thus, it is certainly easy to see why the evidence for an important positive impact of venture capital firms on the birth of biotech firms may have appeared stronger in previous work than seems warranted based on fuller models: Since venture capital firms have developed around a number of great universities, their presence proxies for intellectual human capital in the absence of more direct measures; if they are the only variable indicating presence of great universities and their faculties, they enter positively even if their packaging activities result in a negative direct effect on births.

The decade-lagged dependent variable is added to the full fundamentals model in column (f) of Table 2. Doing so primarily has the effect of weakening the significance of the top-quality universities variable (but, see the annual model below) due to significant multicollinearity between the variables.[13] One interpretation of this positive coefficient on the lagged dependent variable is that agglomeration effects strengthen the impact of fundamentals on regional development. However, the statistical properties of poisson regressions with lagged dependent variables are somewhat problematic so such regressions and their estimated standard errors should be viewed cautiously.

In conclusion, the intellectual human capital variables play a strong role in determining where the U.S. biotech industry developed during the 1980's. We have been able to identify particular star scientists who appear to play a crucial role in the process of spillovers and geographic agglomeration over and above that which would be predicted based on university reputation and scientists supported by federal grants alone. The strong positive role of venture capital variable reported previously is not supported for firm births. Indeed, the data tell us that there were fewer firms founded, other things equal, where there more venture capital firms. It is left to future research to explore whether firms which are associated with particular star scientists or were midwifed by venture capitalists are more successful than other firms.[14]

B. *The Annual Model*

We next report analogous poisson regressions exploiting the panel nature of our data set with observations for the 183 regions for each of the years 1976 through 1989. Tables 3 and 4 report poisson regressions for this entire panel.

Column (a) of Table 3 reports the results using only the counts of stars and their collaborators active each year in each region. As with the long-run models in Table 2, examination of the data suggested that these effects—particularly for stars—were nonlinear so we add squared values in column (b). Again, as the number of stars increases, their marginal contribution diminishes eventually passing through zero.

These nonlinearities might reflect the declining value over time of the intellectual human capital as we have measured it. Basically

[12] This hypothesis was derived from anecdotal evidence, but note that the top nine of Ernst & Young's list of top-ten companies by 1993 market valuation (G. Steven Burrill and Kenneth B. Lee, Jr., 1994 p. 54) were located and founded in regions richly endowed with venture capital firms: Boston (3), San Francisco (3), Los Angeles (1), San Diego (1), and Seattle (1).

[13] In the alternative version of Table 2 (see footnote 9 above), the coefficient on the lagged dependent variable was nearly as large as in Table 2, so the significant positive coefficient does not arise from firms born 1976–1980 appearing in both the current and lagged dependent variables.

[14] See Zucker et al. (1994) for our first effort to the determinants of success of firms after birth.

TABLE 3—ANNUAL POISSON REGRESSIONS ON THE BIRTH OF BIOTECH-USING FIRMS BY REGION AND YEAR, 1976–1989

	(a)	(b)	(c)	(d)	(e)	(f)
Constant	−1.591*** (0.032)	−1.918*** (0.041)	−2.148*** (0.057)	−4.447*** (0.226)	−4.491*** (0.349)	−4.687*** (0.565)
Number stars active in the region and year	0.157*** (0.020)	0.529*** (0.051)	0.270** (0.088)	— —	0.361*** (0.080)	0.282** (0.103)
Number collaborators active in the region and year	0.043*** (0.013)	0.083* (0.035)	0.047 (0.049)	— —	0.013 (0.047)	0.032 (0.052)
(Number stars active in the region and year)²	— —	−0.022*** (0.002)	−0.014* (0.006)	— —	−0.015** (0.005)	−0.014 (0.008)
(Number collaborators active in the region and year)²	— —	−0.001 (0.001)	0.000 (0.001)	— —	0.000 (0.001)	0.001 (0.002)
Number stars active in the region and year × DUMMY1986–89	— —	— —	−0.219 (0.113)	— —	−0.298** (0.102)	−0.245 (0.128)
Number collaborators active in the region and year × DUMMY1986–89	— —	— —	0.117 (0.067)	— —	0.115 (0.064)	0.027 (0.081)
(Number stars active in the region and year × DUMMY1986–89)²	— —	— —	0.006 (0.007)	— —	0.009 (0.006)	0.007 (0.008)
(Number collaborators active in the region and year × DUMMY1986–89)²	— —	— —	−0.001 (0.002)	— —	−0.001 (0.002)	0.001 (0.002)
Number top-quality universities in the region in 1981	— —	— —	0.444*** (0.125)	— —	0.472*** (0.095)	0.462*** (0.109)
Number faculty with federal support in the region in 1979–80	— —	— —	0.625*** (0.093)	— —	0.982*** (0.094)	0.930*** (0.093)
Number venture capital firms in the region and year[a]	— —	— —	— —	0.019** (0.007)	−0.028*** (0.006)	−0.024** (0.008)
Total employment (all industries) in the region and year	— —	— —	— —	0.173*** (0.051)	−0.081 (0.048)	−0.117* (0.055)
Average wages per job in the region and year	— —	— —	— —	0.153*** (0.010)	0.125*** (0.016)	0.132*** (0.017)
Earnings-price ratio (Standard & Poors 500) for year	— —	— —	— —	−0.024 (0.016)	−0.026 (0.026)	−0.017 (0.039)
Number firms active in the region at end of previous year	— —	— —	— —	— —	— —	0.020 (0.013)
Number firms active in all U.S. at end of previous year	— —	— —	— —	— —	— —	−0.000 (0.000)
Births of biotech firms in the region for previous year	— —	— —	— —	— —	— —	0.054 (0.034)
Log-likelihood	−1677.0	−1429.1	−1274.3	−1669.5	−1202.3	−1184.6
Restricted log-likelihood	−2238.5	−2238.5	−2238.5	−2238.5	−2238.5	−2238.5

Notes: $N = 2562$. Standard errors (adjusted by Wooldridge, 1991 Procedure 2.1) are in parentheses below coefficients.

[a] For years after 1981, the number of venture capital firms in a region is held constant at the 1981 level to avert simultaneity problems.

* Significantly different from 0 at the 5-percent level.

** Significantly different from 0 at the 1-percent level.

*** Significantly different from 0 at the 0.1-percent level.

as the knowledge diffuses we expect that more and more stars will result in less and less payoff to any one of them if he or she were to start a firm, and indeed stars are less likely to result in birth of firms after 1985 than before. This is illustrated in column (c) of Table 3 where we add four interaction terms in which these counts and their squares have been multiplied

by a dummy DUMMY1986–89 equal to 1 during 1986–1989 and 0 otherwise, as well as the other intellectual human capital terms. During 1986–1989 the positive effect of stars is sharply reduced while that of collaborators more than triples.[15] Nonetheless, we should view this inference cautiously since the significance values of the interaction terms for stars and collaborators with DUMMY1986–89 fall between 0.10 and 0.05, except for stars in the full fundamentals model in column (e) where the stars interaction term is significant at the 0.01 level.

Thus, we see that (at least during the first decade of this industry) localities with outstanding scientists having the tacit knowledge to practice recombinant DNA were much more likely to see new firms founded and preexisting firms begin to apply biotechnology. There is some evidence that as knowledge about gene splicing diffused and the tacit knowledge lost its scarcity and extraordinary value, the training function of universities became more important relative to the attraction of great scientists to an area. It is interesting that the quadratic term for stars is negative, suggesting diminishing returns (or possibly just proportionately fewer, larger firms) rather than the increasing returns suggested by standard views of knowledge spillovers which posit uninternalized, positive external effects from university scientists.[16] In the same regression in column (c), we see that, beyond the identified stars and collaborators, university quality and federal support are also significant measures of intellectual human capital relevant to firm founding.

Column (d) of Table 3 leads to the same conclusions with panel data as found for the same column in Table 2: The economic variables enter significantly with the expected signs if the intellectual human capital variables are omitted from the regression. However, unlike the previous long-run case, we can now enter the earnings-price ratio.[17] Here this variable enters with the correct sign, but does not even reach the 0.10 level of significance.

Column (e) of Table 3 presents the annual full fundamentals model incorporating the intellectual human capital and other variables. The results for the intellectual human capital measures are robust while the sign of the venture capital variable turns significantly negative as in the long-run model and the employment variable becomes insignificant (and negative).

Column (f) of Table 3, analogously to Table 2, adds a lagged dependent variable to the full fundamentals model. We also included the one-year lagged regional and national counts of firms using biotechnology as dynamic influences reflecting local agglomeration effects and market competition effects, respectively. None of the three dynamic variables enter significantly although their signs are consistent with some geographic agglomeration.

Thus, taken as a whole the results summarized in Table 3 support the strong role of intellectual human capital variables in determining the development of the American biotech industry.

The role of the economic variables, particularly the number of venture capital firms in the region, is explored further in Table 4. The table presents representative results for births in the entrant and incumbent subcomponents of firm entry into biotechnology. We see in columns (a) and (b) that if only the economic variables are introduced we get all the expected signs at appropriate significance [except for employment in (a) and the earnings-price ratio in both], including a result consistent with conventional wisdom that the number of venture capital firms has a sig-

[15] To compute the effects of stars in the 1986–1989 period, we need to add the coefficients of the number of active stars and the coefficient of the same variable interacted with DUMMY1986–89 and then do the same for the two terms involving the squared values of these variables. An analogous approach yields the effect of collaborators during 1986–1989. We examined also interactions with dummy variables for 1976–1980 and with a time trend. Since the coefficients were very small and statistically insignificant for interaction terms involving 1976–1980 dummies, we believe the reported form more accurately reflects the time or diffusion dependence than a negative trend throughout the period.

[16] We are indebted to Jeff Armstrong for this point.

[17] The earnings-price ratio had to be dropped from these analyses because it is available only nationally over time.

TABLE 4—ANNUAL POISSON REGRESSIONS ON THE BIRTH OF BIOTECH-USING ENTRANTS AND INCUMBENTS BY REGION AND YEAR, 1976–1989

	(a) Entrants	(b) Incumbents	(c) Entrants	(d) Incumbents	(e) Entrants	(f) Incumbents
Constant	−4.726***	−5.798***	−4.843***	−5.673***	−4.928***	−5.228***
	(0.284)	(0.563)	(0.409)	(0.902)	(0.669)	(1.285)
Number stars active in the region and year	—	—	0.414***	0.323	0.351**	0.242
	—	—	(0.095)	(0.165)	(0.124)	(0.169)
Number collaborators active in the region and year	—	—	−0.006	0.000	0.012	0.019
	—	—	(0.053)	(0.105)	(0.059)	(0.101)
(Number stars active in the region and year)²	—	—	−0.016**	−0.016*	−0.017	−0.015
	—	—	(0.006)	(0.008)	(0.009)	(0.011)
(Number collaborators active in the region and year)²	—	—	0.001	0.002	0.000	0.001
	—	—	(0.002)	(0.003)	(0.002)	(0.003)
Number stars active in the region and year × DUMMY1986–89	—	—	−0.227*	−0.519*	−0.196	−0.456
	—	—	(0.113)	(0.237)	(0.147)	(0.251)
Number collaborators active in the region and year × DUMMY1986–89	—	—	0.096	0.233	0.011	0.144
	—	—	(0.071)	(0.141)	(0.090)	(0.153)
(Number stars active in the region and year × DUMMY1986–89)²	—	—	0.007	0.018	0.006	0.015
	—	—	(0.007)	(0.010)	(0.010)	(0.013)
(Number collaborators active in the region and year × DUMMY1986–89)²	—	—	−0.001	−0.004	0.001	−0.002
	—	—	(0.002)	(0.003)	(0.003)	(0.004)
Number top-quality universities in the region in 1981	—	—	0.440***	0.479*	0.410**	0.447
	—	—	(0.110)	(0.205)	(0.126)	(0.238)
Number faculty with federal support in the region in 1979–80	—	—	0.973***	1.114***	0.932***	1.041***
	—	—	(0.112)	(0.296)	(0.107)	(0.295)
Number venture capital firms in the region and year[a]	0.023**	0.006	−0.029***	−0.027*	−0.024**	−0.024
	(0.009)	(0.013)	(0.007)	(0.012)	(0.009)	(0.013)
Total employment (all industries) in the region and year	0.128	0.296**	−0.110	−0.052	−0.149*	−0.078
	(0.067)	(0.098)	(0.058)	(0.098)	(0.067)	(0.103)
Average wages per job in the region and year	0.156***	0.139***	0.123***	0.113**	0.127***	0.114**
	(0.012)	(0.024)	(0.018)	(0.039)	(0.020)	(0.040)
Earnings-price ratio (Standard & Poor's 500) for year	−0.036	−0.033	−0.022	−0.056	−0.016	−0.082
	(0.021)	(0.043)	(0.031)	(0.070)	(0.046)	(0.092)
Number firms active in the region at end of previous year	—	—	—	—	0.023	0.024
	—	—	—	—	(0.015)	(0.025)
Number firms active in all U.S. at end of previous year	—	—	—	—	−0.000	−0.001
	—	—	—	—	(0.000)	(0.001)
Share of biotech firms in the region for previous year	—	—	—	—	0.037	0.055
	—	—	—	—	(0.041)	(0.061)
Log-likelihood	−1265.1	−486.3	−945.9	−386.8	−935.8	−382.9
Restricted log-likelihood	−1628.7	−607.9	−1628.7	−607.9	−1628.7	−607.9

Notes: N = 2562. Standard errors (adjusted by Wooldridge, 1991 Procedure 2.1) are in parentheses below coefficients.

[a] For years after 1981, the number of venture capital firms in a region is held constant at the 1981 level to avert simultaneity problems.

* Statistically significantly different from 0 at the 5-percent level.

** Statistically significantly different from 0 at the 1-percent level.

*** Statistically significantly different from 0 at the 0.1-percent level.

cantly positive effect on the birth of new firms but an insignificant effect on the birth of subunits of existing firms which would not normally be financed by venture capital firms. The full fundamentals model is reported in columns (c) and (d) for births of entrants and incumbents, respectively, which is to be compared to column (e) for all firm births in Table 3. Again, in the presence of intellectual human capital the simple economic story does not hold up: the coefficients of venture capital firms and total employment turn negative, significantly so in the former case. Similar results are obtained in the dynamic versions of the full model reported in columns (e) and (f) of Table 4. The robustness of the negative venture capital coefficient remains a puzzle for future work, particularly in view of Yolanda K. Henderson's (1989) evidence that, despite some significant localization, most investments by venture capitalists cross regional boundaries.

IV. Summary and Conclusions

The American biotechnology industry which was essentially nonexistent in 1975 grew to 700 active firms over the next 15 years. In this paper, we show the tight connection between the intellectual human capital created by frontier research and the founding of firms in the industry. At least for this high-tech industry, the growth and location of intellectual human capital was the principal determinant of the growth and location of the industry itself. This industry is a testament to the value of basic scientific research. The number of local venture capital firms, which appears to be a positive determinant when intellectual human capital variables are excluded from the regressions, is found to depress the rate of firm birth in an area, perhaps due to the role of these venture capital firms in packaging a number of scientists into one larger firm which is likely to go public sooner.

We conclude that the growth and diffusion of intellectual human capital was the main determinant of where and when the American biotechnology industry developed. Intellectual human capital tended to flourish around great universities, but the existence of outstanding scientists measured in terms of research productivity played a key role over, above, and separate from the presence of those universities and government research funding to them. We believe that our results provide new insight into the role of research universities and their top scientists as central to the formation of new high-tech industries spawned by scientific breakthroughs. By being able to quantitatively identify individuals with the ability both to invent and to commercialize these breakthroughs, we have developed new specificity for the idea of spillovers and in particular raised the issue of whether spillovers are best viewed as resulting from the nonappropriability of scientific knowledge or from the maximizing behavior of scientists who have the ability to appropriate the commercial fruits of their academic discoveries.

Data Appendix

The data used here are generally in panel form for 14 years (1976–1989) and 183 regions (functional economic areas as defined by the BEA). Frequently, the data are aggregates of data at the zip code or county level. Lagged variables include data for 1975 in the unlagged form. These data sets, part of our ongoing project on "Intellectual Capital, Technology Transfer, and the Organization of Leading-Edge Industries: Biotechnology," will be archived upon completion of the project in the Data Archives at the UCLA Institute for Social Science Research. A full description of the data is available from the authors upon request.

Biotechnology Firms

The starting point for our firm data set covered the industry as of April 1990 and was purchased from NCBC (1991), a private firm which tracks the industry. This data set identified 1075 firms, some of which were duplicates or foreign and others of which had died or merged. Further, there were a significant number of firms missing which had exited prior to April 1990. For these reasons, an extensive effort was made to supplement the NCBC data with information from Bio... (1989–1993) and an industry data set provided by a firm in the industry which was al...

the ancestor of the *Bioscan* data set (Cetus Corp., 1988).

We generally counted entry of firms by adding up for each year and region the number of entrants founded and incumbents first using biotechnology. A few special cases should be noted: Where a firm enters the data set due to the merger of a entrant and another firm, we count it for the purposes of this paper as a continuation of the original entrant and not a new birth (the older entrant if two are involved). If firms already in the data set merge and one continues with the other(s) absorbed, the enterprise is counted as the continuing enterprise and not a new birth.

Scientists

Star scientists and their collaborators were identified as described in the text. Individual scientists are linked to locations through the institutional affiliations reported in their publications in the article data set. The discovery of genetic sequences is recognized by *GenBank*'s assignment to an article of a "primary accession number" to identify each. The 12 additional stars added to the 315 with more than 40 primary accession numbers thus had 20 or more articles with at least one primary accession number and 20–40 primary accession numbers total.

Articles

Our article data set consists of all 4,061 articles in major journals listed in *GenBank* as reporting genetic-sequence discoveries for which one or more of our 327 stars were listed as authors. (A small number of unpublished papers and articles appearing in proceedings volumes and obscure journals were excluded to permit the hand coding detailed below.) All of these articles were assigned unique article ID numbers and collected by hand. For each article, scientist ID numbers are used to identify the order of authorship and the institutional affiliation and location for each author on each article. This hand coding was necessary because, under the authorship traditions in these fields, the head of the laboratory who is often the most prestigious author frequently

authors on 18.3 percent of the articles and last authors on 69.1 percent of the 4,031 articles remaining after excluding the 30 sole-authored articles.[18] Unfortunately, only first- and/or corresponding-author affiliations are available in machine-readable sources.[19]

The resulting authorship data file contains 19,346 observations, approximately 4.8 authors for each of the 4,061 published articles. Each authorship observation gives the article ID number, the order of authorship, the scientist ID number of one of our stars and collaborators, and an institutional ID number for the author's affiliation which links him or her to a particular institution with a known zip code as of the publication date of the article.

Citations

We have collected data for 1982, 1987, and 1992 on the total number of citations to each of our 4,061 published articles listed in the Institute for Scientific Information's *Science Citation Index* (1982, 1987, 1992). These citation counts are linked to the article and authorship data set by the article ID number. The citations were collected for articles if and only if they appeared in the article data set; so scientists are credited with citations only insofar as they are to the 4,061 articles reporting genetic-sequence discoveries and published in major journals.

Universities

Our university data set consists of all U.S. institutions listed as granting the Ph.D. degree in any field in the Higher Education General Information Survey (HEGIS), Institutional Characteristics, 1983–1984 (U.S. Department of Education, National Center for Education

[18] This positional tradition holds across national boundaries: As a percentage of articles coauthored by their fellow nationals, American stars are 16.4 percent of first authors and 71.2 percent of last authors, compared to 21.2 percent and 63.1 percent, respectively, for Japanese, and 19.7 percent and 69.2 percent for other nationalities.

[19] The *Science Citation Index* lists up to six of the af-

Statistics, 1985). Each university is assigned an institutional ID number, a university flag, and located by zip code based on the HEGIS address file. Additional information described in the text was collected from Jones et al. (1982) for those universities granting the Ph.D. degree in biochemistry, cellular/molecular biology, and/or microbiology which we define as *"biotech-relevant"* fields.

Research Institutes and Hospitals

For those U.S. research institutions and hospitals listed as affiliations in the article data set, we assigned an institutional ID number and an institute/hospital flag, and obtained an address including a zip code as required for geocoding. No additional information has been collected on these institutions.

Venture Capital Firms

We created a venture capital firm data set by extracting from the Pratt (1982) directory the name, type, location, year of founding, and interest in funding biotech firms. This information was extracted for all venture capital which were legally permitted to finance start-ups. This latter requirement eliminated a number of firms which are chartered under government programs targeted at small and minority businesses. This approach accounts includes founding date of firms appearing in the 1982 Pratt directory, excluding those firms that may have either entered thereafter or existed in earlier years but exited before the directory was compiled.

Other Economic Variables

Total employment and average earnings per job by region and year are as reported by the Bureau of Economic Analysis based on county level data in U.S. Department of Commerce (1992b): Total employment is from Table K, line 010 (in millions of persons). Average earnings is from Table V, line 290 (wage & salary disbursements, other labor income, and proprietors income per-job in thousands of current dollars), deflated by the implicit price deflator for personal consumption expenditures. The annual data for the implicit price deflator for personal consumption expenditures were taken from U.S. Department of Commerce (199..., p. 247, line 16) as updated in the July 19... *Survey of Current Business*, (p. 92, line 1...). The S&P 500 earnings-price ratio was taken from *CITIBASE* (1993), series FSEXP.

REFERENCES

Arrow, Kenneth J. "Economic Welfare and the Allocation of Resources for Invention," in Richard R. Nelson, ed., *The rate and direction of inventive activity: Economic and social factors*. National Bureau of Economic Research Special Conference Series, Vol. 13. Princeton, NJ: Princeton University Press, 1962, pp. 609–25.

Bioscan, Vols. 3–7. Phoenix, AZ: Oryx Press, 1989–1993.

Burrill, G. Steven and Lee, Kenneth B., Jr. *Biotech 94: Long-term value, short-term hurdles*. Ernst & Young's Eighth Annual Report on the Biotechnology Industry, San Francisco: Ernst & Young, 1994.

Cameron, A. Colin and Trivedi, Pravin K. "Regression-Based Tests for Overdispersion in the Poisson Model." *Journal of Econometrics*, December 1990, *46*(3), pp. 347–64.

Cetus Corp. "Biotechnology Company Data Base," predecessor source for *Bioscan*. Computer printout, Cetus Corp., 1988.

CITIBASE: Citibank Economic Database. Machine-readable database, 1946–June 199... New York: Citibank, 1993.

GenBank, Release 65.0. Machine-readable database. Palo Alto, CA: IntelliGenetics, Inc., 1990.

Greene, William H. *LIMDEP: User's manual and reference guide, version 6.0*. Bellport, NY: Econometric Software, 1992.

Griliches, Zvi. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, October 1957, *25*(4), pp. 501–22.

______. "The Search for R&D Spillovers." *Scandinavian Journal of Economics*, 199..., Supplement, *94*, pp. 29–47.

Grossman, Gene M. and Helpman, Elhanan. *Innovation and growth in the global economy*. Cambridge, MA: MIT Press, 1991.

Hausman, Jerry; Hall, Bronwyn H. and Griliches, Zvi. "Econometric Models for Count Data with an Application to the Patents-R&D Relationship." *Econometrica*, July 1984, *52*(4), pp. 909–38.

Henderson, Yolanda K. "Venture Capital and Economic Development." Paper presented to the New England Advisory Council, Federal Reserve Bank of Boston, Boston, MA, July 11, 1989.

Institute for Scientific Information. *Science Citation Index*, ISI compact disc editions, machine-readable database. Philadelphia: Institute for Scientific Information, 1982, 1987, 1992.

Jaffe, Adam B. "Real Effects of Academic Research." *American Economic Review*, December 1989, *79*(5), pp. 957–70.

Jaffe, Adam B.; Trajtenberg, Manuel and Henderson, Rebecca. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *Quarterly Journal of Economics*, August 1993, *63*(3), pp. 577–98.

Jones, Lyle V.; Lindzey, Gardner and Coggeshall, Porter E., eds. *An assessment of research-doctorate programs in the United States: Biological sciences*. Washington, DC: National Academy Press, 1982.

Kenney, Martin. *Biotechnology: The university-industrial complex*. New Haven, CT: Yale University Press, 1986.

Lerner, Joshua. "Venture Capitalists and the Decision to Go Public." *Journal of Financial Economics*, June 1994, *35*(3), pp. 293–316.

______. "Venture Capitalists and the Oversight of Private Firms." *Journal of Finance*, March 1995, *50*(1), pp. 301–18.

Mansfield, Edwin. "Academic Research Underlying Industrial Innovations: Sources, Characteristics, and Financing." *Review of Economics and Statistics*, February 1995, *77*(1), pp. 55–65.

Nelson, Richard R. "The Simple Economics of Basic Scientific Research." *Journal of Political Economy*, June 1959, *67*(3), pp. 297–306.

Nelson, Richard R. and Romer, Paul M. "Science, Economic Growth, and Public Policy," in Bruce L. R. Smith and Claude E. Barfield, eds., *Technology, R&D, and the economy*. Washington, DC: Brookings Institution and American Enterprise Institute, 1996, pp. 49–74.

Nelson, Richard R. and Winter, Sidney G. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press, 1982.

North Carolina Biotechnology Center. *North Carolina Biotechnology Center U.S. companies database*, machine-readable database. Research Triangle Park, NC: North Carolina Biotechnology Center, April 16, 1992.

Pratt, Stanley E. *Guide to venture capital sources*, 6th Ed. Englewood Cliffs, NJ: Prentice-Hall, 1982.

Romer, Paul M. "Increasing Returns and Long-Run Growth." *Journal of Political Economy*, October 1986, *94*(5), pp. 1002–37.

______. "Endogenous Technological Change." *Journal of Political Economy*, October 1990, Part 2, *98*(5), pp. S71–S102.

Rosenberg, Nathan. *Inside the black box: Technology and economics*. Cambridge: Cambridge University Press, 1982.

U.S. Department of Commerce, Bureau of Economic Analysis, Economics and Statistics Administration. *National income and product accounts of the United States, volume 2, 1959–88*. Washington, DC: U.S. Government Printing Office, 1992a.

______. *Regional economic information system, version 1.3*, CD-ROM, machine-readable database. Washington, DC: Bureau of Economic Analysis, May 5, 1992b.

U.S. Department of Education, National Center for Education Statistics. *Higher education general information survey (HEGIS), institutional characteristics, 1983–84*, machine-readable database, ICPSR 8291. Ann Arbor, MI: Inter-University Consortium for Political and Social Research, circa 1985.

Wooldridge, Jeffrey M. "On the Application of Robust, Regression-Based Diagnostics to Models of Conditional Means and Conditional Variances." *Journal of Econometrics*, January 1991, *47*(1), pp. 5–46.

Zucker, Lynne G.; Brewer, Marilynn B.; Oliver, Amalya and Liebeskind, Julia. "Basic Science as Intellectual Capital in Firms: Information Dilemmas in rDNA Biotechnology

Research." Working paper, UCLA Institute for Social Science Research, 1993.

Zucker, Lynne G. and Darby, Michael R. "Star Scientists and Institutional Transformation: Patterns of Invention and Innovation in the Formation of the Biotechnology Industry." *Proceedings of the National Academy of Sciences*, November 12, 1996, *93*(23), pp. 12709–12716.

Zucker, Lynne G.; Darby, Michael R. and Armstrong, Jeff. "Intellectual Capital and the Firm: The Technology of Geographically Localized Knowledge Spillovers." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 4946, December 1994.

Zucker, Lynne G.; Darby, Michael R. and Brewer, Marilynn B. "Intellectual Capital and the Birth of U.S. Biotechnology Enterprises." National Bureau of Economic Research (Cambridge, MA) Working Paper No. 4653, February 1994.

Sunk Costs and Firm Value Variability: Theory and Evidence

By VAL EUGENE LAMBSON AND FARRELL E. JENSEN*

Dynamic competitive models are interesting tools for studying industry behavior over time. Although there is some overlap, such models of industry evolution can be divided into two classes: learning (or internal shocks) models and external shocks models.

Learning models focus on idiosyncratic shocks that are internal to the firm. For example, Boyan Jovanovic's (1982) passive learning model posits that firms naturally acquire noisy information about their efficiency as they produce. Firms that receive too many unfavorable signals conclude they are inefficient, are scrapped, and are replaced by new entrants. Richard Ericson's and Ariel Pakes' (1995) active learning model focuses on research and development components of industry evolution. Firms make investment decisions knowing that their future productivity is a stochastic function of their current productivity and their current level of investment. Firms that fall too far behind in the technology race are scrapped and are replaced by new entrants.

Unlike learning models—see also Hugo A. Hopenhayn (1992)—external shocks models emphasize common shocks faced by all firms. Examples of external shocks include demand shocks, factor price shocks, and exchange rate fluctuations. If firms are similar, as assumed by Avinash Dixit (1989) and Lambson (1992), then they will be similarly affected. If firms differ, as in Lambson's (1991) model of endogenous firm heterogeneity, then they will be affected differently.

The various models of industry evolution are not logically inconsistent, rather they focus on different phenomena. The resulting theoretical insights generate several predictions that a small body of empirical work on industry dynamics has addressed. For example, Timothy Dunne et al. (1989) document the importance of entry and exit while pointing out some intriguing empirical regularities that seem consistent with the dynamic competitive approach. Pakes and Ericson (1998) contrast some empirical implications of passive and active learning models and conclude that passive learning is more important in retailing whereas active learning is more important in manufacturing. Lambson and Jensen (1995) argue that agricultural data are consistent with external shocks models.

This paper reviews the predictions of dynamic competitive models regarding the variability of firm value and subjects them to empirical testing. Learning models predict that intra-industry firm value variability will be greater in higher sunk cost industries. External shocks models predict that intertemporal firm value variability will be greater for firms in higher sunk cost industries.

The empirical results are consistent with both the learning models and the external shocks models. Specifically, stock market prices and publicly available debt positions provide estimates of firm value while the book value of property, plant, and equipment (capital costs) serves as a proxy for firm-level sunk costs. Regressions of firm value variability (whether intra-industry or intertemporal) on industry averages of capital costs yield significantly positive coefficients. Controlling for firm size does not alter the conclusions.

Section I reviews the implications of dynamic competitive models pertaining to firm value variability. Section II outlines the empirical framework and describes the data. Section III contains the empirical

* Department of Economics, Brigham Young University, Provo, UT 84602. We are grateful to the Silver Fund for financing the acquisition of the COMPUSAT data and [illegible] Larsen for helping us access it. We also thank [illegible] Kretsch, Julie Elston, Mark Roberts, and two anonymous referees for their suggestions, as well as [illegible]

I. Some Implications of Dynamic Competitive Models

Formal descriptions of various dynamic competitive models are in the papers cited in the introduction; here a less formal discussion will suffice. Dynamic competitive models endow expected-present-value-maximizing firms with infinite horizons and place them in an uncertain environment where some actions entail sunk costs. Specifically, at the beginning of each of countably infinitely many periods, firms observe current conditions and then make their entry and exit decisions. A new firm enters if its expected present value of being in business (given an optimal exit rule) exceeds its entry cost. Similarly, an existing firm exits if its expected present value of remaining in business (and optimally exiting later) falls below its scrap value. Firm i's entry cost and scrap value will be denoted E_i and S_i, respectively. Their difference, $E_i - S_i$, is firm i's sunk cost.

After firm i enters, it is buffeted by events that affect its profitability. These events are characterized by the three stochastic processes—γ_i, φ, and y—described below. The t^{th} period realizations of γ_i, φ, and y are denoted γ_{it}, φ_t, and y_t, respectively.

The random variable γ_{it}, emphasized by learning models, is idiosyncratic to firm i. It incorporates internal shocks and any other heterogeneity. Jovanovic (1982) would interpret γ_{it} as a random signal that depends positively on firm i's exogenous and unobserved efficiency: firms with higher γ_{it} values believe they are more efficient and thus have higher expected values. Ericson and Pakes (1995) would interpret γ_{it} to be firm i's efficiency as a function of past R&D successes: the current level of efficiency γ_{it} is a stochastic function of the previous level of efficiency $\gamma_{i,t-1}$ and the previous R&D investment decision.

The random variable φ_t, emphasized by external shocks models, is common to all firms. Dixit (1989) would interpret φ_t as factor and output prices, which evolve exogenously. Lambson (1991, 1992) would interpret φ_t as factor prices, among other things, but endogenizes output prices.[1]

Finally, the random variable y_t describes the number of existing firms and their characteristics after entry and exit decisions have been implemented in period t. The value of y_t depends on y_{t-1} and on firms' reactions to the [illegible] and γ_{it} realizations.

Let $V_{it}(\gamma_{it}, \varphi_t, y_t)$ denote the (expected present) value of firm i at time t. The formal definition of V_{it} differs across the various models, but in each case it can be derived from the solution to a dynamic programming problem that specifies the firm's optimal production and exit rules given the stochastic outcomes and the equilibrium behavior of the other firms. Dynamic competitive models make predictions about both intra-industry and intertemporal firm value variability. Intuitively, the basic insight is that entry and exit put upper and lower bounds on firm values: high firm values tend to provoke entry and low firm values tend to provoke exit. Where sunk costs are larger, however, firm values can rise more without provoking entry and fall more without provoking exit. Thus, the range of firm values should be larger where sunk costs are higher.

Predictions about intra-industry firm value variability arise from differences in γ_{it} across firms, and hence are naturally associated with learning models. Let I_t be the set of firms in industry I at time t. To begin simply, suppose that all firms in and potential entrants to industry I have the same entry cost $\bar{E}_I$ and the same scrap value $\bar{S}_I$. Further assume that the firms with the highest values can be rapidly imitated by entrants—so that $\max_{i \in I_t} V_{it} = \bar{E}_I$—and that the firms with the lowest values are indifferent between exiting and remaining in business—so that $\min_{i \in I_t} V_{it} = \bar{S}_I$. Under these assumptions, the range of firm values in industry I at time t—defined by $R_i(I, t) = \max_{i \in I_t} V_{it} - \min_{i \in I_t} V_{it}$—satisfies

$$R_i(I, t) = \bar{E}_I - \bar{S}_I. \tag{1}$$

[1] Here there is some overlap with learning models. Lambson (1991) would interpret γ_i as a constant that indexes firm i's technology choice. Firms that enter with different technologies are affected differently by changes in φ.

Equation (1) predicts that regressing intra-industry firm value variability $R_i(I, t)$ on the industry's sunk cost $\bar{E}_I - \bar{S}_I$ will yield a coefficient of one.

Unfortunately, the assumptions underlying equation (1) are stringent ones. To see the effects of relaxing them, reinterpret $\bar{E}_I$ and $\bar{S}_I$ as the average entry cost and average scrap value, respectively, of incumbents in industry I. Add and subtract industry I's average sunk cost, $\bar{E}_I - \bar{S}_I$, to rewrite $R_i(I, t) = \max_{i \in I_t} V_{it} - \min_{i \in I_t} V_{it}$ as

$$(2) \quad R_i(I, t) = (\max_{i \in I_t} V_{it} - \bar{E}_I) + (\bar{E}_I - \bar{S}_I) - (\min_{i \in I_t} V_{it} - \bar{S}_I).$$

The first right-hand-side term is how much the most successful firm's value exceeds the average entry cost. The second term is the average sunk cost for existing firms. The third term reflects how much the least valuable firm's value exceeds the average scrap value.

Given the assumptions underlying equation (1), the first and third right-hand-side terms of equation (2) are zero and equation (2) reduces to equation (1). More generally, the first term may be of either sign. (It will be positive if times are good and the most fortunate incumbent enjoys inimitable advantages due to lucky R&D outcomes, access to unique factors, or because potential entrants have high entry costs. It will be negative if times are so bad that even the most successful firms have low values.) Similarly, the third term may be of either sign. (It will be positive, for example, if all firms have the same scrap value and the least valuable firm strictly prefers to remain in business. It will be negative, for example, if the least valuable firm has a lower than average scrap value and is close to indifferent between exiting and remaining.) As long as the first and third terms are uncorrelated with the second term, however, equation (2) predicts that the regression of $R_i(I, t)$ on $\bar{E}_I - \bar{S}_I$ will yield a positive coefficient: the omitted variables $\max_{i \in I_t} V_{it} - \bar{E}_I$ and $\min_{i \in I_t} V_{it} - \bar{S}_I$ will be swept into the constant and the error term. Thus, under these assumptions, the theory predicts a higher intra-industry range of firm values in higher sunk cost industries.

Predictions about intertemporal firm value variability arise from differences in φ_t over time and hence are naturally associated with external shocks models. Let $R_t(i)$ be the range of firm i's value over time, that is, $R_t(i) = \max_t V_{it} - \min_t V_{it}$. In the absence of intra-industry heterogeneity—that is, when γ_{it} is the same for all $i \in I_t$ and, furthermore, all firms in and potential entrants to industry I have the same entry cost $\bar{E}_I$ and the same scrap value $\bar{S}_I$—equilibrium requires

$$(3) \quad \bar{E}_I \geq V_{it} \geq \bar{S}_I$$

for all $i \in I_t$ and for all t because entry prevents V_{it} from exceeding the entry cost and exit prevents V_{it} from falling below the scrap value. If firm i is observed long enough so that there are both entry-provoking good times (when $V_{it} = \bar{E}_I$) and exit-provoking bad times (when $V_{it} = \bar{S}_I$), then expression (3) implies

$$(4) \quad R_t(i) = \bar{E}_I - \bar{S}_I.$$

Specifically, regressing intertemporal firm value variability $R_t(i)$ on the industry's sunk cost $\bar{E}_I - \bar{S}_I$ should yield a coefficient of one.

Intra-industry heterogeneity—reflected in different γ_i processes, different realizations of a given γ_i process or differences in entry costs or scrap values—weakens the predictions of equation (4). (For example, the values of inimitably efficient firms may rise above average entry costs and fail to fall to average scrap values.) To allow for heterogeneity, reinterpret $\bar{E}_I$ and $\bar{S}_I$ as the average entry cost and average scrap value, respectively, of incumbents in industry I. Add and subtract industry I's average sunk cost, $\bar{E}_I - \bar{S}_I$, to rewrite $R_t(i) = \max_t V_{it} - \min_t V_{it}$ as

$$(5) \quad R_t(i) = (\max_t V_{it} - \bar{E}_I) + (\bar{E}_I - \bar{S}_I) - (\min_t V_{it} - \bar{S}_I).$$

In the homogeneous case the first and third right-hand-side terms are zero, as argued above. More generally, if intra-industry heterogeneity is additive over time, so that a more (respectively, less) efficient firm is always more (less) valuable by a constant than a

numeraire firm whose value fluctuates between $\bar{E}_t$ and $\bar{S}_t$, then the first and third terms cancel and the results are the same as for the homogeneous case. Unfortunately, this additivity assumption is still very stringent. However, as long as the first and third terms are uncorrelated with the second term, regressing $R_t(i)$ on $\bar{E}_t - \bar{S}_t$ should yield a positive coefficient: the omitted terms will simply sweep into the constant and the error term.

In summary, under the assumptions, dynamic models of industry evolution predict that intra-industry firm value variability is higher in high sunk cost industries and that intertemporal firm value variability is higher for firms in high sunk cost industries.

II. The Empirical Framework and the Data

Direct tests of the empirical implications discussed in Section I require data on sunk costs and firm values. Unfortunately, detailed data on sunk costs seem to be unavailable and detailed data on firm values are difficult to obtain except for publicly traded firms. However, by focusing on established firms and using a proxy for sunk costs, one can construct simple tests of the implications of dynamic competitive models for firm value variability.

The data used in the study are from four files of Standard and Poor's Compustat Services: the Primary, Supplementary, and Tertiary Industrial Files and the Industrial Research File. The Primary Industrial File has information on a sample of companies from the S&P Industrial, Utilities, Transportation, and Financial Indexes. The Supplementary Industrial File includes information on additional companies listed on major stock exchanges. The Tertiary File contains information on banks, life insurance companies, railroads, property and casualty insurance companies, real estate investment trusts, and additional companies from the S&P Utilities, Transportation, and Financial Indexes. Finally, the Industrial Research File contains companies that have been deleted from other Compustat files because of acquisition or merger, bankruptcy, leveraged buyout, reverse acquisition, or because the company no longer files with the SEC. The data, reported annually, cover the 20-year period 1973–1992 for 4,534 firms. The largest firm had an average value over the sample period of 363 billion dollars while the smallest firm had an average value over the sample period of 2.8 million dollars. The highest reported average book value of property, plant, and equipment was 118 billion dollars while the smallest average book value was reported to be zero.

The dependent variables suggested by the models are the intra-industry and intertemporal ranges of real firm value. Firm value (in millions of dollars) was defined as the real market value of firm equity (that is, the fiscal year-end price of a share of common stock times the number of shares outstanding at the end of the fiscal year) plus the real book value of total liabilities, where real values were calculated using the GNP deflator with 1987 as the base year. (Although it would have been preferable to have the market value of the liabilities rather than the book value, such data were unavailable.) The intra-industry range of firm value for a given year was defined as the difference between the highest firm value and the lowest firm value within a four-digit SIC code. The intertemporal range of a firm's value was defined as the difference between the maximum and minimum values observed for the firm between 1973 and 1992 inclusive. Observations for which the range was trivially zero were discarded. (For the intra-industry range these correspond to industries for which only one firm was reported for the year. For the intertemporal range these correspond to firms that were in the sample for only one year.)

The measure of the intra-industry range of firm values will understate its theoretical counterpart to the extent that extreme-valued firms (e.g., new firms) are not publicly traded or are not in the sample for other reasons. Furthermore, the four-digit SIC codes may not perfectly coincide with the theoretical notion of an industry. Even so, the measure will be a good proxy for its theoretical counterpart to the extent that the biases are not systematic.

The measure of the intertemporal range of firm value almost certainly understates its theoretical counterpart because the assumption that the observed firms have enjoyed both good times and bad times (with values close to their upper and lower extremes) during

TABLE 1—SUNK COST REGRESSIONS

Model	Variable	Coefficient	Standard error	R^2	Observations
Learning	Intercept	1,958.37	211.76	0.34	6,266
	K	9.71	0.17		
External shocks	Intercept	1,625.21	138.56	0.04	4,209
	K	1.21	0.09		

Notes: For the learning model an observation corresponds to an industry in a given year, the dependent variable is the intra-industry range of real firm value and K is the intra-industry mean of firm-level capital costs. For the external shocks model an observation corresponds to a firm, the dependent variable is the intertemporal range of real firm value, and K is the intra-industry mean of the intertemporal mean of firm-level capital costs.

sample period may not be satisfied, especially for firms that are only in the sample for a short time. Nevertheless, it should be positively related to its theoretical counterpart and the underestimate should become smaller as the length of time a firm is observed increases.

The independent variable suggested by the theory is the average sunk cost in each industry. Unfortunately, such costs are difficult to measure. We used the real gross book value (in millions of dollars) of property, plant, and equipment (capital costs) as a proxy for firm-level sunk costs, where real values were calculated each year by using the GNP deflator with 1987 as the base year. In the learning models context, we used the intra-industry average of this proxy as a proxy for the average sunk cost in each industry.

In the intertemporal context—in contrast to the simplifying theoretical assumption that sunk costs are stable over time—the data report that firms' capital costs change over time as firms build new plants, retool old plants, and remove scrapped equipment from their books. We "smoothed" the data by using the average capital costs for each firm over time as a proxy for firm-level sunk costs. We then used the intra-industry average of this proxy as a proxy for the average sunk cost in each industry.

On one hand, capital costs probably overstate sunk costs in two ways: they ignore depreciation and they ignore scrap values. Capital that depreciates away is no longer sunk, but real depreciation is difficult to measure because available estimates are based on accounting definitions that have little economic meaning. Scrap values are even harder to measure: one must know what the equipment and other embodiments of sunk costs would be worth outside the industry. On the other hand, capital costs probably understate sunk costs in other ways: other costs than capital costs—such as the opportunity cost of the entrepreneurial time used to start a company, legal fees, etc.—are also sunk. We do not claim that capital costs are a good measure of sunk costs, only that, in the absence of good firm-level data on sunk costs, they are a good proxy: other things equal, firms with higher sunk costs probably have higher capital costs and the converse.

III. Empirical Results

Learning models predict that regressing the intra-industry range of firm value at a given time on a proxy for the industry's average sunk cost should yield a positive coefficient. External shocks models predict that regressing the intertemporal range of a firm's value on a proxy for its industry's average sunk cost should yield a positive coefficient. The results of these regressions are in Table 1 and are consistent with the theory: the coefficient in each case is positive and highly significant.

These results, however, do not distinguish the sunk cost theory of firm value variability from other conceivable size-related theories. Specifically, suppose that production in industry I requires a sunk cost of σ_I per worker. If L_i is the number of workers employed by firm i then firm i's sunk cost is $E_i - S_i = L_i\sigma_I$ and, other things equal, is higher for larger firms. This suggests an alternative interpretation of

TABLE 2—SIZE REGRESSIONS

Model	Variable	Coefficient	Standard error	R^2	Obs[illegible]
Learning	Intercept	1,120.42	226.12	0.31	[illegible]
	Range (L)	140.99	2.68	—	
External shocks	Intercept	731.88	115.30	0.35	[illegible]
	L	168.08	3.70	—	

Notes: For the learning model an observation corresponds to an industry in a given year, the dependent vari[illegible] intra-industry range of real firm value and Range (L) is the intra-industry range of firm-level employment. For the [illegible] shocks model an observation corresponds to a firm, the dependent variable is the intertemporal range of real fir[illegible] and L is the intertemporal mean of firm-level employment.

TABLE 3—NORMALIZED SUNK COST AND SIZE REGRESSIONS

Model	Variable	Coefficient	Standard error	R^2	Observ[illegible]
Learning	Intercept	−2,648.91	1,213.04	0.15	[illegible]
	K/L	94.28	2.86	—	
	Range (L)	−9.95	14.12	—	
External shocks	Intercept	−685.75	679.23	0.07	[illegible]
	K/L	11.56	0.69	—	
	L	−0.26	20.94	—	

Notes: For the learning model an observation corresponds to an industry in a given year, the dependent variable [illegible] intra-industry range of real firm value per worker, K/L is the intra-industry average of firm-level capital costs per [illegible] and Range (L) is the intra-industry range of firm-level employment. For the external shocks model an obs[illegible] corresponds to a firm, the dependent variable is the intertemporal range of real firm value per worker, K/L is the [illegible] industry mean of the intertemporal mean of firm-level capital costs per worker, and L is the intertemporal mean of [illegible] level employment.

Table 1. Table 2 shows, not surprisingly, that intra-industry firm value variability is positively related to intra-industry firm size variability and that intertemporal firm value variability is positively related to firm size (where firm size is measured by employment). Thus Table 1 does not rule out the possibility that sunk cost is merely a proxy for firm size variability (in the learning context) or firm size (in the external shocks context) which, in turn, are related to the true determinants of firm value variability.

Fortunately, the sunk cost theory of firm value variability can be distinguished from other size-related theories because the predictions do not change when one controls for size: sunk costs can be normalized by setting $L_i = 1$ in the previous paragraph and focusing on sunk costs and firm value per worker.[2] [illegible] associated normalized regressions [illegible] strongly positive sunk cost coefficients [illegible] this method of controlling for firm size is [illegible] arguably inadequate. Since higher sunk [illegible] industries exhibit higher firm size vari[illegible] in the data, the sunk cost variable may sti[illegible] a proxy for firm size variability in the [illegible] malized learning regression. Similarly, [illegible] large firms are more capital intensive in [illegible] data, the sunk cost variable may still [illegible] proxy for firm size in the normalized ex[illegible] shocks regression. For these reasons, we [illegible]

[2] This is analogous to our previous work on agri[illegible] Lambson and Jensen (1995)—which can be [illegible] as normalizing with respect to land inputs.

intra-industry firm size variability and firm size (as measured by the intra-industry range of firm-level employment and firm-level employment, respectively) directly to the respective normalized regressions. As shown in Table 3, the sunk cost coefficients remain strongly positive whereas the size coefficients (in contrast to Table 2) become insignificant. This suggests that firm value variability is explained by sunk costs rather than other size related variables.

As further evidence, note that the theory predicts that the ranking of industries by their intra-industry firm value variability and the ranking of industries by the average intertemporal firm value variability of their firms should be similar. Consistent with this, the Spearman correlation coefficient for the two rankings (of 334 industries) is 0.88.

The reported regressions are very simple. It is important to emphasize that the theory suggests that they should be: the theory leads one to focus on sunk costs as a major determinant of the range of firm value. Whether it is caused by output demand shocks, input price shocks, changes in regulatory regimes, random outcomes of research and development, or anything else, the range of firm value is constrained by the difference between entry costs and scrap values. Hence, a simple regression of the range of firm value on a measure of sunk costs is precisely what the theory calls for when these variables can be accurately measured.

We conclude that sunk costs are an important determinant of firm value variability, as is predicted by dynamic competitive models of industry evolution.

REFERENCES

Dixit, Avinash. "Entry and Exit Decisions Under Uncertainty." *Journal of Political Economy*, June 1989, *97*(3), pp. 620–38.

Dunne, Timothy; Roberts, Mark J. and Samuelson, Larry. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics*, November 1989, *104*(4), pp. 671–98.

Ericson, Richard and Pakes, Ariel. "Markov-Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, January 1995, *62*(1), pp. 53–82.

Hopenhayn, Hugo A. "Entry, Exit, and Firm Dynamics in Long Run Equilibrium." *Econometrica*, September 1992, *60*(5), pp. 1127–50.

Jovanovic, Boyan. "Selection and the Evolution of Industry." *Econometrica*, May 1982, *50*(3), pp. 649–70.

Lambson, Val Eugene. "Industry Evolution with Sunk Costs and Uncertain Market Conditions." *International Journal of Industrial Organization*, June 1991, *9*(2), pp. 171–96.

______. "Competitive Profits in the Long Run." *Review of Economic Studies*, January 1992, *59*(1), pp. 125–42.

Lambson, Val Eugene and Jensen, Farrell E. "Sunk Costs and the Variability of Firm Value Over Time." *Review of Economics and Statistics*, August 1995, *77*(3), pp. 535–44.

Pakes, Ariel and Ericson, Richard. "Empirical Implications of Alternative Models of Firm Dynamics." *Journal of Economic Theory*, 1998 (forthcoming).

Measuring Consumer Surplus with Unknown Hicksian Demands

By IAN J. IRVINE AND WILLIAM A. SIMS *

A standard problem in welfare analysis is that estimated demand functions, or labor supply functions, which are sufficiently flexible to capture the variation in behavior observed in data, frequently cannot easily be integrated back to an explicit cost or utility function, despite satisfying the integrability conditions.[1] Some functions are easily integrable to yield closed forms, but impose unreasonable restrictions on the data. The consequent trade-off between the generality of the econometric functions and the possibility of obtaining the corresponding cost or utility function has been examined exhaustively by Nicholas Stern (1986), in the context of labor supply.

Recent practice, when a closed-form utility or expenditure function cannot easily be obtained from estimated, but well-behaved, demand functions,[2] is to use numerical methods to approximate welfare measures (Yrjö Vartia, 1983; Kathy Hayes and Susan Porter-Hudak, 1987; Breslaw and Smith, 1995; Hausman and Whitney Newey, 1995). Because of the inherent complexity of such methods, however, various single-step procedures have also been employed. One such measure is the Marshallian consumer surplus, which can be used to approximate either the compensating or equivalent variation. An alternative approach involves employing a Taylor approximation of either the indirect utility function (George McKenzie and Ivor Pearce, 1976) or the cost function (Andreu Mas-Colell et al., 1995 p. 89). The latter technique, assuming a second-order expansion, amounts to using the tangent to the Hicksian demand at initial prices as an approximation to the true, but unknown, Hicksian demand when calculating the welfare change. Mas-Colell et al. (1995 p. 90) suggest that this linear approximation to the Hicksian demand is both simple to apply and, in the case of small price changes, more accurate than the Marshallian consumer surplus as a measure of welfare change.

In this paper we propose an alternative two-step procedure, which is at least as simple as the two previously mentioned one-step approximations (i.e., the linear Hicksian approximation and the Marshallian consumer surplus), but is generally more accurate. The procedure is based upon what is termed the Slutsky compensated demand (or supply in the labor case), involves no numerical integration and requires only an understanding of elementary utility theory.

In Section I we illustrate the idea by means of a geometric example. In Section II we apply the method to a specific well-behaved labor supply function where the cost and utility functions are recoverable only by means of numerical expansions. We then examine accuracy by considering cases where the cost function *is* easily obtainable. In Section III we develop the theoretical properties of our measure. We show that it is locally path independent and develop results on relative errors which support the findings from the numerical examples.

* Department of Economics, Concordia University, Montreal, Canada, H3G 1M8. Irvine would like to thank the University of Sydney for the use of research facilities while working on this paper.

[1] The integrability problem involves the conditions under which we can be sure that observable demands are the result of a utility-maximizing process. The sufficient conditions are called the integrability conditions and are developed in Leonid Hurwicz and Hirofumi Uzawa (1971). In essence they are the conditions for a well-behaved demand system: that the demand functions are homogeneous of degree zero, add up, and have symmetric, negative semidefinite substitution terms, in addition to certain regularity conditions. Even where these conditions are fulfilled, it may be very difficult to find closed forms for cost and utility functions. The technique suggested in this paper can be used in such cases.

[2] Jerry Hausman (1981 p. 666) shows, for a single price change, "... that for many applications no approximation is needed." But as is pointed out in Jon Breslaw and J. Barry Smith (1995 p. 313), "... since this involves solving differential equations, Hausman's technique becomes difficult to implement when the demand functions are complex or when there is more than one price change."

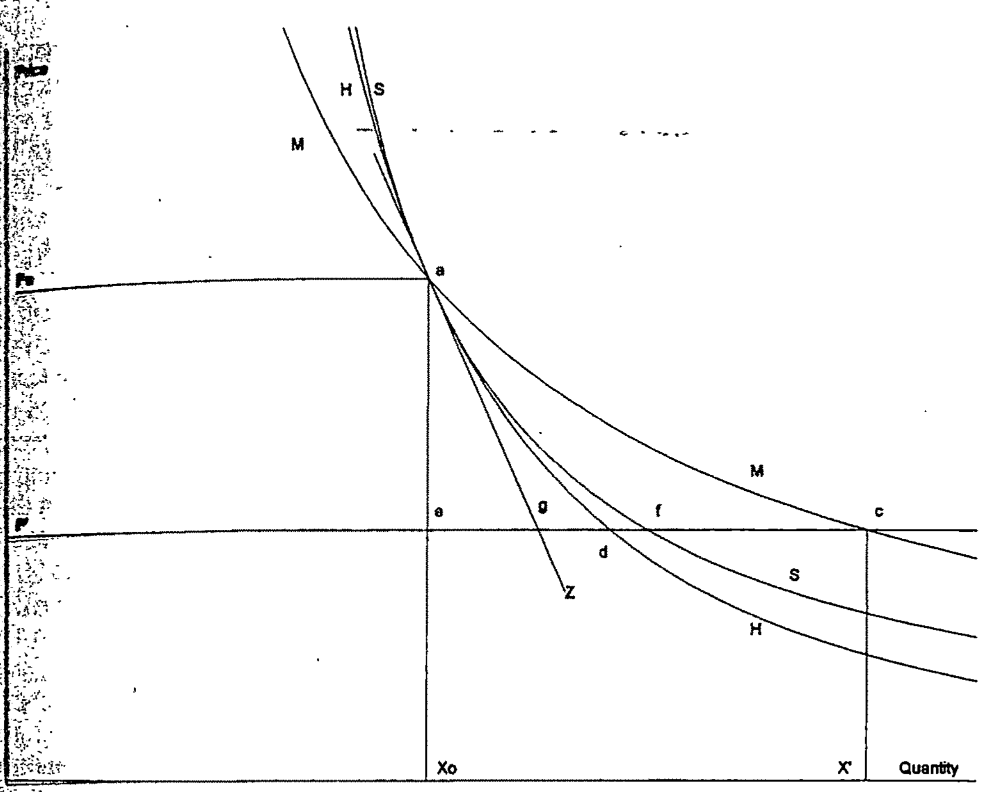


FIGURE 1. THE SLUTSKY DEMAND

[illegible], in Section IV, we relate our approach [illegible] of Hausman and Newey (1995).

I. Slutsky-Based Compensation

In some microeconomics texts (e.g., Milton [illegible], 1962, or Eugene Silberberg, 1990) [illegible] exercise is to examine the relationship between Hicksian and Slutsky compensated demand curves. Figure 1 is typically [illegible] such exercises. We shall use the term [illegible] compensation to denote the income [illegible] required so that a consumer can [illegible] the bundle of goods purchased prior to [illegible] change. The amount of a good purchased [illegible] a given set of prices with such com[illegible] we will denote by $x^s = x^s(\mathbf{p}, \mathbf{x}^0)$ [illegible] is the vector of goods which the con[illegible] at the initial price vector $\mathbf{p}^0$. In [illegible] Hicks compensation is the amount re[illegible] a particular level of indiffer[illegible] associated demand is $x^H = x^H(\mathbf{p},$ [illegible] $(\mathbf{x}^0))$ is the initial level of utility [illegible] The Marshallian demand will be denoted by $x^M = x^M(\mathbf{p}, y)$ where y is income. The Slutsky demand is derived by substituting the vector product $\mathbf{px}^0$ for y in the Marshallian demand. For a normal good, x, the relationship between these demands (Jacob Mosak, 1942) is that the Slutsky (S) is above, but tangent to the Hicksian (H) demand at the point where they intersect the Marshallian (M) demand[3]—point a in Figure 1.[4]

When the Hicksian demand is known, the area to the left of the Hicksian demand

[3] If x were inferior the Marshallian demand would be the steepest of the three demands and the Slutsky demand curve would lie everywhere beneath the Hicksian curve, except at the point of tangency. A technical derivation of the relationship between Hicksian and Slutsky demands is provided in the Appendix at the end of this paper.

[4] The demands in Figure 1 come from a Cobb-Douglas utility function $u = x_1 x_2$ with a budget constraint $x_1 + x_2 = 8$. The proximity of x^S to x^H is thus not simply constructed to support our approach. All figures are developed with *Maple V R3* using exact functions, and are presented to scale.

represents the monetary value to the consumer of a price change. When this demand is defined for the initial level of utility, u^0, the resulting measure is the compensating variation, CV, and when it is defined for the new (post-price change) utility level it is the equivalent variation, EV. When the Hicksian demand is unknown an approximation is required. Since the Slutsky demand is always obtainable from the Marshallian demand, our proposal is to use the Slutsky demand to approximate these measures. In the case of a single price change this amounts to computing the area p^0afp' in Figure 1 as an approximation of the unknown area p^0adp' when price falls from p^0 to p'.[5]

To examine the relationship between the Slutsky-based measure of the CV and other methods, it is convenient to define the CV mathematically. Letting e denote the expenditure function, the CV is given by

$$(1) \qquad CV = e(\mathbf{p}^0, u^0) - e(\mathbf{p}', u^0),$$

where $\mathbf{p}$ is a vector of prices at the initial ($\mathbf{p}^0$) and final ($\mathbf{p}'$) equilibria. Expanding $e(\mathbf{p}', u^0)$ around the initial price and utility combination by means of a Taylor series, and considering only one price change, we obtain

$$(2) \qquad e(\mathbf{p}', u^0) = e(\mathbf{p}^0, u^0) + \frac{\partial e(\mathbf{p}^0, u^0)}{\partial p}\Delta + 0.5\frac{\partial^2 e(\mathbf{p}^0, u^0)}{\partial p^2}\Delta^2 + R_2,$$

where R_2 is the remainder term in the series, and Δ is the price change $(p' - p^0)$. If the quadratic terms alone form a good approximation, then

$$(3) \qquad CV \cong -x^H(\mathbf{p}^0, u^0)\Delta - 0.5\frac{\partial x^H(\mathbf{p}^0, u^0)}{\partial p}\Delta^2$$

since the derivative of the expenditure function is the Hicksian demand. $\partial x^H(\mathbf{p}^0, u^0)/\partial p$ can be evaluated by the Slutsky equa[tion for] small changes in p. That is:

$$(4) \qquad \frac{\partial x^H(\mathbf{p}^0, u^0)}{\partial p} = \frac{\partial x^M(\mathbf{p}^0, y^0)}{\partial p} + x^H\frac{\partial x^M(\mathbf{p}^0, y^0)}{\partial y}$$

and this gradient is therefore obtain[able] from a knowledge of the parameters [of the] Marshallian demand function. This app[roach] then, which is essentially what is presen[ted in] Mas-Colell et al. (1995 p. 89), amounts to [us]ing the tangent (Z) to the Hicksian dem[and in] Figure 1 at the initial equilibrium and [approx]imating the area p^0adp' by p^0agp'.[6]

II. Application

A desirable characteristic of labor [supply] functions is that they display flexibility [over] different ranges of the wage w and unearned income m. This is what makes the linear la[bor] supply model of limited value for tax r[eform] analysis—particularly at the extremes of [the] wage and unearned income values (M[ervyn] King, 1987). We define the underlying d[irect] utility function as $u(c, L)$, and the co[st, or] expenditure, function as $m = c - wL$, [where] L is labor supplied and c is consumption.

A. A Quadratic Labor Supply Functi[on]

Consider now a labor supply function, [from] Stern (1986), which is quadratic in wages, [w,] and unearned income, m.

$$(5) \qquad L = \alpha w + \beta m + \lambda w^2 + \mu m^2 + \nu w m + \gamma.$$

This function permits L to be backward be[nd]ing at high wage rates—if $\lambda < 0$. It also all[ows] $\partial^2 L/\partial m^2 < 0$ (e.g., more BMWs make [work] progressively less attractive). The sta[...]

[5] To avoid repetition, we develop the theory using the CV only, although John Kay (1980) has noted the advantage of the EV over the CV.

[6] The third one-step procedure, Marshallian co[nsumer] surplus, is measured as the area to the left of [the] Marshallian demand (M). In Figure 1 this is rep[resented] by the area p^0acp'.

approach to recovering the utility function from estimated demands or supplies is first to obtain the expenditure function, then the indirect utility function by inversion. The direct utility function (if required) is obtained by differentiating the indirect utility with respect to the normalized price vector, and substituting the prices. Following Hurwicz and Uzawa (1971), the expenditure function can be obtained (where possible) by integrating the partial differential equation

$$(6) \qquad \frac{\partial m}{\partial w} = -L,$$

where m is the expenditure function. For $\mu = 0$ in (5) this is feasible, and yields an indirect utility (Stern, 1986 p. 175)

$$(7) \qquad v(w, m) = m e^{\beta w + 0.5 \nu w^2} - Q(w)$$

where,

$$Q(w) = \int \{(-\alpha w - \lambda w^2 - \gamma) e^{\beta w + 0.5 \nu w^2}\} \, dw.$$

When μ above is nonzero, recoverability becomes complex as we examine below. However, the Slutsky-compensated labor supply, L^S, can be obtained easily from (5) by substituting for m using the relation $m = c^0 - wL^0$. This yields

$$(8) \qquad L^S = w(\alpha - \beta L^0 - 2\mu c^0 L^0 + \nu c^0) + w^2(\lambda + \mu L^{0^2} - \nu L^0) + (\beta c^0 + \mu c^{0^2} + \gamma) = a_0 + a_1 w + a_2 w^2.$$

Equation (8) is plotted in Figure 2 for a given set of coefficients.[7] In order to estimate the welfare cost of a wage reduction, brought about by a tax rate change, it is straightforward to integrate (8) with respect to w to obtain the answer. For example, suppose the wage rate is reduced by 20 percent as a result of a tax, then the area to the left of L^S in Figure 2 between $(w = 4.15)$ and $(w = 3.32)$ is a measure of the welfare loss. Essentially L^S is approximating the unknown Hicksian labor supply function.

What makes the Slutsky demand tractable is that we do not have to integrate a differential equation of the form (6) with a nonlinear term in m present. It is the quadrature of the labor supply equation in m which makes the integration to a cost function complex. We should emphasize that it is possible to obtain a cost and indirect utility function for the quadratic function given above by numerical methods, as shown by Hausman (1981 Appendix). These functions are given, respectively, by

$$(9) \qquad e(w, \bar{u}) = \frac{(\beta + \nu w)\nu^2}{2\mu} - \frac{1}{\mu} \frac{\tilde{W}_1' + \bar{u} \tilde{W}_2'}{\tilde{W}_1 + \bar{u} \tilde{W}_2}$$

and

$$(10) \qquad v(w, m) = \frac{h\tilde{W}_1 - \tilde{W}_1'}{\tilde{W}_2' - h\tilde{W}_2},$$

where $h = -\mu m + \nu(\beta + \nu w)^2/2$, and the $\tilde{W}$'s are obtained by means of infinite expansions involving the coefficients in (5) and prices. These expansions are defined in, for example, Milton Abramowitz and Irene Stegun (1970 p. 686). Accordingly, the *CV* or *EV* can be computed by using the expenditure function, evaluated at different values of the price vector by using a computer algorithm as described by Hausman. Of course we cannot derive a Hicksian labor supply by the application of Shephard's lemma to (9) without the use of numerical methods.

The conclusion we draw from this is that the exact recoverability of the cost and utility functions in this limited case of a simple quadratic function with *only one price change* is

[7] The set of values chosen are such as to make the function ... initial equilibrium, comparable to Hausman's ... labor supply example. The equilibrium ... given $w = 4.15$, and $m = 8{,}236$ is $L = 1{,}790.2$. ... which yield this are: $\alpha = 1{,}500.0$, $\beta = -0.01$, $\lambda = -90.0$, $\mu = -.000002$, $\gamma = -1{,}596.5$, $\nu = -0.0313$. This set of values satisfies the integrability conditions.

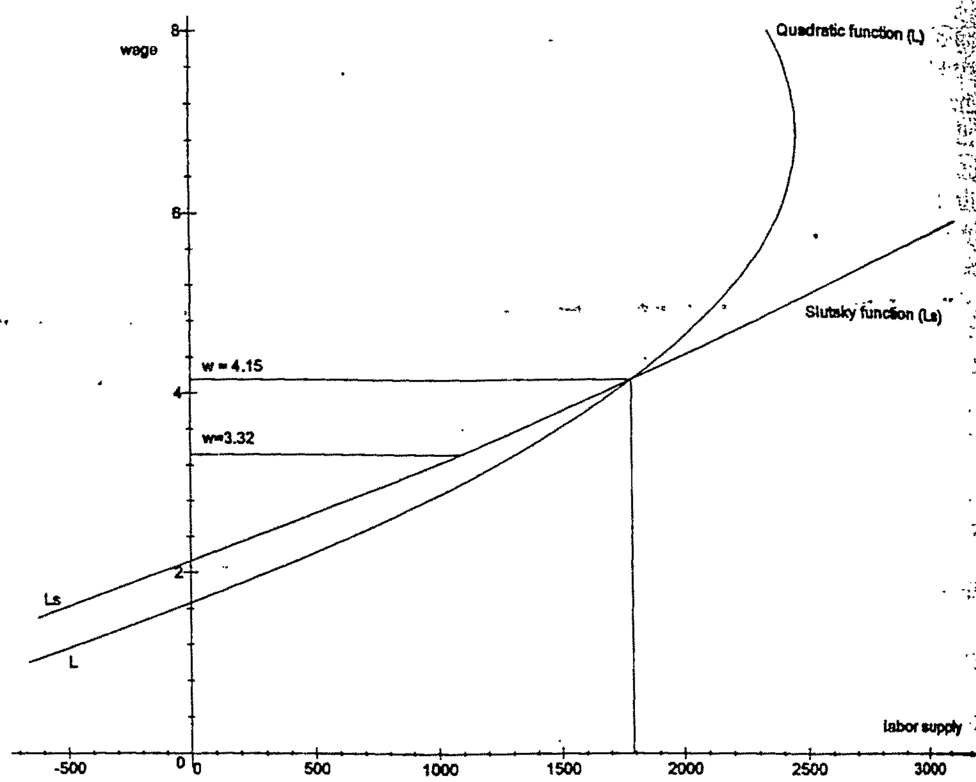


FIGURE 2. THE QUADRATIC LABOR SUPPLY EXAMPLE

not at all trivial and cannot be obtained exactly without infinite expansions. Hausman has been quite clear on this, though he has frequently been misinterpreted. The fact that the expenditure and utility functions can be written as above should not mask the fact that infinite expansions underlie the coefficients, meaning that closed-form analytic solutions do not support (9) and (10).

Correspondingly, if we wished to impose a cubic form on the data, or wanted to change more than one price in the system, or wanted to use expansions of finite order, the recoverability of the cost and utility functions becomes yet more complex or less exact.

This then leads to the question of how good different methods are at approximating the exact measure. To test the exactness of the Slutsky demand or supply equation in computing the CV or EV we take two simple examples in which the analytic cost functions, and therefore an exact welfare measure, are available.

B. *Accuracy*

Consider first the linear labor supply function used by Hausman (1980, 1981):

$$L = \alpha w + \delta m + s. \tag{11}$$

The associated indirect utility function is

$$v(w, m) = e^{\delta w}\left(m + \frac{\alpha}{\delta} w + \frac{s}{\delta} - \frac{\alpha}{\delta^2}\right), \tag{12}$$

where e is the exponential operator. The resulting Hicksian supply function L^H is

$$L^H = \frac{\alpha}{\delta} + \delta e^{-\delta w} u, \tag{13}$$

where u can be either the initial (u^0) or (u') level of utility attained. Finally, Slutsky labor supply, L^S, is obtained

substituting for m in (11) using the relation $m = c^0 - wL^0$

$$(14)\quad L^s = w(\alpha - \delta L^0) + \delta c^0 + s.$$

For the set of values given in Hausman (1980, 1981)[8] the true CV is found to be $-\$1{,}246.65$ and the estimate of the CV based on the Slutsky demand is $-\$1{,}238.3$, an error of about one-half of 1 percent. In contrast, the estimate using the Marshallian function is $-\$1{,}315.4$, an error of 6 percent.[9]

As a second example we computed the CV associated with a 50-percent increase in p_1 in the example underlying Figure 1, given in footnote 4. We found the Slutsky error was 0.72 percent, the error associated with using the tangency to the Hicksian demand was -2.67 percent, and the error associated with the Marshallian demand was -9.79 percent.

III. Theoretical Properties

In this section, we compare the Slutsky welfare approximation with the other two one-step approximations. First the theoretical relationship between the errors from measuring CV with the Slutsky and Marshallian demands is developed in a manner which does not depend on the functional forms of the demands. Following this we compare the Slutsky measure with that based on the second-order approximation to the expenditure function. In addition, the theoretical path independence of the Slutsky approach is investigated.

A. A Willig-Type Result

Robert Willig (1976) derived a measure of the error associated with the use of a Marshallian, rather than a Hicksian, demand to estimate the CV or EV. Willig's result can be illustrated, following Robin Boadway and Neil Bruce (1984 p. 218), with the use of Figure 1. A fall in price from p^0 to p' yields a welfare gain, as measured by the CV, equal to the area p^0adp'. The Marshallian approximation to this is p^0acp', yielding an error acd. Geometrically this can be approximated by $\frac{1}{2}dc\Delta p$[10] and an expression for this is easily derived: dc is the income effect associated with an income change of CV $(=p^0adp')$. That is, $dc \simeq \partial x/\partial y \Delta y$, where $\Delta y = CV$. Accordingly the Marshallian error is given by

$$(15)\qquad adc \simeq \frac{1}{2}\eta\frac{x}{y}(CV)\Delta p,$$

where η is the income elasticity of demand.

The error associated with using the Slutsky demand is given by adf. As before we can derive an expression for the distance df: The movement from c to f is attributable to the Slutsky compensation—in this case negative. The Slutsky compensation is defined by $x^0\Delta p$ $(=p^0aep')$ in the case of a single good. It therefore follows that the Slutsky income effect, df, is due simply to an increase in income defined by the area aed. Accordingly the error associated with using the Slutsky demand is

$$(16)\qquad adf \simeq \frac{1}{2}\eta\frac{x}{y}(aed)\Delta p.$$

The area aed can be interpreted as a gain from the elimination of a deadweight loss if the price fall resulted from eliminating a tax equal to Δp. Referring to aed as the deadweight loss, DWL, we therefore have the result that the ratio of the Slutsky error to the Marshallian error is approximately DWL/CV.

[8] m is set at \$8,236, w^0 is 4.15, w^1 is 3.32. The estimated parameters are $\alpha = 495.1$, $\delta = -0.125$, $s = 765.1$. These values yield utility of $u^0 = -27{,}387.2$, and $u^1 = -27{,}210.4$.

[9] It is of historical interest to note that Hausman's original estimate of the CV incorporates a numerical error. His estimate of the CV is $-\$2{,}056$ and on the basis of this he imputes an error of 44 percent to the Marshallian estimate. Thus, as a measure of the true welfare loss, the Marshallian measure is not as faulty as Hausman proposes and his objection against using Willig's approach is not justified. Nevertheless, *as a measure of the pure component of the CV*, the Marshallian-based measure differs significantly from the true measure. Using Hausman's (1981) gasoline example, Irvine and Sims show that the Marshallian demand-based estimate results in an error in excess of 22 percent, whereas the Slutsky measure results in an error of less than 1.1

[10] Making triangular assumptions is equivalent to using the linear terms of a Taylor-series expansion.

A corresponding result can be derived for a price increase or for the EV. The magnitude of the ratio, for a given initial consumption, depends upon the price elasticity of the Hicksian demand. Specifically, the smaller its price elasticity the greater is the error from using the Marshallian demand. In the limit, if the good in question is perfectly complementary with the aggregate of other goods, the Slutsky and Hicksian demands coincide and DWL, as measured by the Hicksian demand, tends to zero.

While these results are remarkably straightforward, they illustrate that there is no need to use Marshallian demands to estimate CV or EV, when the Hicksian demand cannot be obtained. If the Marshallian demand can be integrated, so can the Slutsky demand, and it yields an estimate of the welfare change which is an order of magnitude smaller than that which comes from the Marshallian demand.

B. *The Second-Order Approximation to an Expenditure Function*

The other simple one-step approximation to welfare changes considered uses a second-order approximation to an expenditure function at initial prices. This, in essence, involves using the linear function, Z, in Figure 1 as an approximation to the Hicksian demand. There are a number of reasons to expect this measure to underperform the Slutsky approximation.[11] Mas-Colell et al. (1995 p. 90) point out that in cases where the price change is large, it is impossible to be sure that the linear Hicksian approximation is better even than the Marshallian consumer surplus approximation. This problem arises because as prices change nothing guarantees that this measure will be sensitive to demand behavior. This is obviously not so with respect to the measure
on either the Marshallian or Slutsky de
And of course, as was shown in the p
section, the Slutsky measure performs
than the Marshallian measure, globally

C. *Path Independence*

An important property of Hicksian
mands, which is of special concern whe
termining the welfare effects of multiple
changes, is path independence.[12] This c
acteristic of demand systems requires
metric cross-partial price derivatives
$\partial p_j = \partial x_j^H / \partial p_i$. Slutsky demands possess
property locally. That is, at the conditi
price vector $\mathbf{p}^0$, $\partial x_i^S / \partial p_j = \partial x_j^S / \partial p_i$ be
the *functions* x^H and x^S are tangent.[13] W
path independence thus holds only locally
should be noted that Marshallian dem
generally do not have this property
locally.

IV. Conclusion

Two final points should be noted. The
is that standard errors for our measures of C
or EV can be obtained readily. This is
from equation (8), since the standard error
the coefficients a_i are known from their eco
metric estimates.

The second point is motivated by Haus
and Newey's recent contribution (1995
They propose that a nonparametric speci
tion for the price and income portion of
mand function will generally yield
reliable estimates than the imposition of
pler forms. They develop a method of e
mating both a value for welfare change
associated standard error. If we adopt
Slutsky demand as a tool of analysis,
polynomial of any order in both prices
income can be fitted to the data. An esti
of welfare change can then be derived by

[11] In the specific cases that we have explored, the Cobb-Douglas utility function (see Section II, subsection B), the indirect addilog from McKenzie and Pearce (1976), and the case of quasi-linear preferences, the Slutsky measure dominated the linear Hicksian approximation, for large and small price changes. In cases in which the Marshallian demand is linear, such as for example Hausman's (1981) gasoline example and his labor supply function (see Section II, subsection B), the linear Hicksian and Slutsky approximations are identical.

[12] An example approximating the CV resulting
multiple price changes, using the Slutsky dema
tions, is provided by Irvine and Sims (1995).

[13] From the equality of their first derivatives
follows that the linear Taylor-series expansions
are equal.

relation $m = \mathbf{px}^0$ in the case of a demand function, or $m = c^0 - wL^0$ in the labor supply function. This can be obtained by proceeding as in the quadratic labor supply example given above, regardless of the order of the polynomials used in estimation, provided the integrability conditions are satisfied by the coefficient estimates.

We conclude by emphasizing that the purpose of this paper is to propose an intuitive, computationally simple and accurate measure of welfare change in cases where estimated demand and supply functions do not yield an easily recoverable cost or utility function. We make no claim that our approach is superior to any of the iterative numerical methods proposed in the literature in recent years. However, in our experimentation with examples where Hicksian functions *can* be obtained from estimated functions, we find that the error associated with using the Slutsky demand to measure the true *CV* or *EV* yields exceedingly small errors for relatively large price changes.

The reason for this accuracy is that the Slutsky demand obviously provides an excellent approximation to the Hicksian demand, as illustrated by their tangency at an equilibrium. It is this tangency which yields local path independence for the Slutsky function and also means that the income effects which make the Marshallian demand less suitable for welfare analysis do not contaminate the Slutsky-based computation in the same way.

APPENDIX: THE SLUTSKY DEMAND

The Slutsky demand can be derived by solving the following problem:

$$\text{(A1)}\quad \text{Maximize } u(\mathbf{x}) \text{ subject to the vector product } \mathbf{px}^0 = \mathbf{px},$$

where $\mathbf{x}$ is a vector of commodities, $\mathbf{p}$ are the corresponding prices, $\mathbf{x}^0$ is the bundle of goods on which the demand is conditioned, and u is a strictly quasi-concave utility function. The Slutsky demand takes the general form

$$\text{(A2)}\quad x_i = x_i^S(\mathbf{p}, \mathbf{x}^0).$$

It is also clear that the following identity holds:

$$\text{(A3)}\quad x_i^S(\mathbf{p}, \mathbf{x}^0) \equiv x_i^M\left(\mathbf{p}, \sum_i p_i x_i^0\right) \equiv x_i^H\left(\mathbf{p}, v\left(\mathbf{p}, \sum_i p_i x_i^0\right)\right),$$

where v is the indirect utility function, x_i^M is the Marshallian demand, and x_i^H is the Hicksian demand. Differentiating (A3) with respect to p_i, and applying Roy's identity,

$$\text{(A4)}\quad \frac{\partial x_i^S}{\partial p_i} = \frac{\partial x_i^H}{\partial p_i} + (x_i^0 - x_i^M)\frac{\partial x_i^M}{\partial y},$$

where x_i^M is the demand for x_i when $y = \Sigma_i\ p_i x_i^0$ and $\partial x_i^H / \partial p_i$ is the slope of the Hicksian demand, x_i^H, conditioned on $u(\mathbf{x}^0)$, at any p.

Two results follow from this when comparing the Hicksian demand conditioned on $u(\mathbf{x}^0)$ and the Slutsky demand conditioned on $\mathbf{x}^0$:

(i) When prices and income are such that the quantity actually demanded (x_i^M) is equal to the bundle on which the Slutsky demand is conditioned (x_i^0), then the Hicksian and the Slutsky demands have the same slope. That is, they are tangent as is shown in Figure 1.

(ii) When prices and income are such that $x_i^M > x_i^0$, then the Hicksian demand is steeper than the Slutsky demand at x_i^M, assuming that x_i is a normal good. Conversely, when prices and income are such that $x_i^M < x_i^0$, then the Hicksian demand is flatter than the Slutsky demand at x_i^M, assuming that x_i is a normal good.

This demonstrates that any particular Hicksian demand is an envelope of Slutsky demands, since this result can be demonstrated for *any* arbitrary bundle, $\mathbf{x}'$, that corresponds to a point on the Hicksian demand conditioned on $u(\mathbf{x}^0)$.

If (A3) is differentiated with respect to p_j, the resulting equation demonstrates the local path independence property discussed in Section III, subsection C.

REFERENCES

Abramowitz, Milton and Stegun, Irene. *Handbook of mathematical functions*. New York: Dover Publications, 1970.

Boadway, Robin and Bruce, Neil. *Welfare economics*. New York: Blackwell, 1984.

Breslaw, Jon and Smith, J. Barry. "A Simple and Efficient Method for Estimating the Magnitude and Precision of Welfare Changes." *Journal of Applied Econometrics*, July–September 1995, *10*(1), pp. 313–27.

Friedman, Milton. *Price theory: A provisional text*. Chicago: Aldine, 1962.

Hausman, Jerry. "The Effects of Wages, Taxes, and Fixed Costs on Women's Labor Force Participation." *Journal of Public Economics*, October 1980, *14*(2), pp. 161–94.

———. "Exact Consumer Surplus and Deadweight Loss." *American Economic Review*, September 1981, *71*(4), pp. 662–76.

Hausman, Jerry and Newey, Whitney. "Nonparametric Estimation of Exact Consumer Surplus and Deadweight Loss." *Econometrica*, November 1995, *63*(6), pp. 1445–76.

Hayes, Kathy and Porter-Hudak, Susan. "Deadweight Loss: Theoretical Size Relationship and the Precision of Measurement." *Journal of Business and Economic Statistics*, January 1987, *5*(1), pp. 47–52.

Hurwicz, Leonid and Uzawa, Hirofumi. "On the Integrability of Demand Functions," in John Chipman, Leonid Hurwicz, Marcel Richter, and Hugo Sonnenschein, eds., *Preferences, utility and demand*. New York: Harcourt Brace Jovanovich, 1971, pp. 114–48.

Irvine, Ian and Sims, William A. "Welfare Measurement when Hicksian Demands are Unknown." Working Paper No. 9513, Concordia University, Montreal, Canada, …

Kay, John. "The Deadweight Loss from a … System." *Journal of Public Economics*, February 1980, *13*(1), pp. 111–20.

King, Mervyn. "The Empirical Analysis of … Reforms," in Truman Bewley, ed., *Advances in econometrics*. Cambridge, U.K.: Cambridge University Press, 1987, pp. …–90.

Mas-Colell, Andreu; Whinston, Michael D. and Green, Jerry. *Microeconomic theory*. New York: Oxford University Press, 1995.

McKenzie, George and Pearce, Ivor. "Exact Measures of Welfare and the Cost of Living." *Review of Economic Studies*, October 1976, *43*(3), pp. 465–68.

Mosak, Jacob. "On the Interpretation of the Fundamental Equation of Value Theory," in Oskar Lange, Francis McIntyre, and Theodore Yntema, eds., *Studies in mathematical economics and econometrics*. Chicago: University of Chicago Press, 1942, pp. 69–74.

Silberberg, Eugene. *The structure of economics: A mathematical analysis*. New York: McGraw-Hill, 1990.

Stern, Nicholas. "On the Specification of Labour Supply Functions," in Richard Blundell and Ian Walker, eds., *Unemployment, search and labor supply*. Cambridge, MA: Cambridge University Press, 1986, pp. 143–…

Vartia, Yrjö. "Efficient Methods of Measuring Welfare Change and Compensated Income in Terms of Ordinary Demand Functions." *Econometrica*, January 1983, *51*(1), pp. 79–98.

Willig, Robert. "Consumer's Surplus Without Apology." *American Economic Review*, September 1976, *66*(4), pp. 589–97.